Bioinformatic Analysis of Differentially-Expressed Genes Corregulated by NOTCH in Breast Cancer

By

Wasima Taslim
20146089

A thesis submitted to the School of Pharmacy in partial fulfillment of the requirements for the degree of Bachelor of Pharmacy (Hons)

School of Pharmacy
BRAC University
May 2024

Declaration

It is hereby declared that

1. The thesis submitted is my/our own original work while completing degree at Brac University.
2. The thesis does not contain material previously published or written by a third party, except where this is appropriately cited through full and accurate referencing.
3. The thesis does not contain material which has been accepted, or submitted, for any other degree or diploma at a university or other institution.
4. I have acknowledged all main sources of help.

Student's Full Name & Signature:

_____

Wasima Taslim
Student ID: 20146089

Approval

The project titled "Bioinformatic Analysis of Differentially-Expressed Genes Corregulated by NOTCH in Breast Cancer" submitted by Wasima Taslim (ID 20146089) of Spring, 2020 has been accepted as satisfactory in partial fulfillment of the requirement for the degree of Bachelor of Pharmacy (Hons.) on 2/5/24.

Examining Committee:

Supervisor:

(Member)

Dr. Muhammad Asaduzzaman
Professor
School of Pharmacy
BRAC University

Approved by:

Program Director:
(Member)

Professor Dr. Hasina Yasmin
Assistant Dean and Program Director
School of Pharmacy
BRAC University

Dean:

Professor Dr. Eva Rahman Kabir
Dean
School of Pharmacy
BRAC University

Ethics Statement

The study does not involve any form of animal trial or clinical trial on humans.

Abstract

Breast cancer development and progression require the involvement of an array of cell signaling pathways and biological processes, the NOTCH signaling pathway being of particular interest. The NOTCH pathway, although extensively regulated, is a highly resourceful pathway due to its apparent relativity with other cellular pathways. This expands the pathway's ability to directly or indirectly influence and regulate other biological processes in order to repress proliferation, angiogenesis, metastasis, differentiation, of malignant cells while promoting their apoptosis in breast cancer. Recently, researchers have focused on NOTCH signaling and its relationship to anti-tumor immunity. Hence, this study intended to conduct a bioinformatic analysis of NOTCH pathway, its related components and genes in breast cancer to identify possible druggable targets (genes) for developing novel anti-cancer molecules in management and treatment of breast cancer. The potential targets include *TEK, ENG, KDR, NOS3, IGF1, PPARG, ACVRL1, ADIPOQ, and TGFBR2*. The current study collected primary data from KEGG Database, GSEA, UALCAN, cBioPortal, GEPIA 2, ShinyGO and SNPs3D to establish a list of genes that are plausible candidates to consider for rational drug development.

Keywords: NOTCH signaling pathway; breast cancer; repress; angiogenesis; apoptosis; anti-tumor immunity; KEGG Database; UALCAN; SNPs3D

Dedication

Dedicated to my parents and every hopeful individual battling breast cancer.

Acknowledgement

I wish to start by thanking Almighty Allah for his boundless grace and blessings, which have steered me along a path of endurance, optimism, and clarity all the while assisting me in overcoming moments of despair, turmoil, and uncertainty during the course of this journey.

Nevertheless, the culmination of this work would be unimaginable without the constant guidance and support of my supervisor, Professor Dr. Muhammad Asaduzzaman, whose intellect, wisdom and professional judgement made the entire process easier for me. It would have been inarguably tough to chose a complicated topic and work on it without his expertise and vast knowledge. I am eternally grateful to him for sparing his valuable time to help me get through every step of this project and helping me recognize and correct my mistakes. He has inspired me to navigate complex topics in the research field and taught me new concepts throughout the entirety of this study.

My greatest appreciation towards Bristy Saha, M.Pharm student, Department of Clinical Pharmacy and Pharmacology, Dhaka University, for mentoring me throughout my thesis work.

I would also like to extend my profound appreciation to our respected Dean, Professor Dr. Eva Rahman Kabir and Deputy Chair and Program Coordinator, Professor Dr. Hasina Yasmin and other faculty members of School of Pharmacy, BRAC University, for creating a challenging yet encouraging educational atmosphere that makes it possible us to explore novel territories and achieve the heights of our abilities.

Lastly, I want to convey my gratefulness to the unwavering source of encouragement in my life, my family, to whom I shall be forever indebted.

Table of Contents

List of Tables

List of Figures

List of Acronyms

ADAM10: ADAM metallopeptidase domain 10

AMPK: AMP-activated protein kinase

BCL2: B-cell lymphoma 2

cAMP: Cyclic adenosine 3′,5′-monophosphate

CC: Correlatiob Coefficient

CsL: cellulose synthase-like

EGF: Epidermal growth factor

ER: Estrogen Receptor

ErbB: Erythroblastic oncogene B

FoxO: Forkhead box O

GLOBOCAN: Global Cancer Observatory

HER2: Human epidermal growth factor receptor 2

HIF-1: Hypoxia-inducible factor 1

Hippo: Salvador-Warts-Hippo

HRT: Hormone replacement therapy

JAK-STAT: Janus kinase/signal transducers and activators of transcription

MAPK: Mitogen-activated protein kinase

MET: Mesenchymal Epithelial Transition

mTOR: Mammalian target of rapamycin

MYC: Master Regulator of Cell Cycle Entry and Proliferative Metabolism

NF-kappa B: Nuclear factor kappa B

NOTCH1: Neurogenic locus notch homolog protein 1

NOTCH2: Neurogenic locus notch homolog protein 2

NOTCH3: Neurogenic locus notch homolog protein 3

NOTCH4: Neurogenic locus notch homolog protein 4

P53: Transformation-related protein 53

PI3K-Akt: Phosphoinositide-3-kinase–protein kinase B/Akt

PKC: Protein kinase C

PR: Progesterone Receptor

Rap1: Ras-proximate-1 or Ras-related protein 1

Ras: Rat sarcoma

SNAIL: Zinc finger protein SNAI1

SVM: Support vector machine

TGF-beta: Transforming growth factor-beta

TNF: Tumor Necrosis Factor

TWIST: Helix-loop-helix protein 38 (bHLHa38) H-twist

VEGF: Vascular endothelial growth factor

Wnt: Wingless-related integration site

Zeb: Zinc Finger E-Box Binding Homeobox

Chapter 1

Introduction

1.1 Breast Cancer

Following decades of both scientific and clinical studies, in addition to trials of potential novel medications, cancer continues to be one of the leading causes of illness and mortality. Cancer has been ranked as the dominant cause of premature death in people under the age of 70. With 19.9 million new cases in, a total of 9.7 million cancer deaths were recorded in 2022 (*Figure 1.1*). Asia reported one-half of all cases and 56.1% of cancer deaths (GLOBOCAN, 2022). The incidence of cancer is about 19% higher in men than in women, although the percentage is not remotely the same in every region. Among women, breast cancer accounts for 1 in 4 cancer cases and 1 in 6 cancer deaths (GLOBOCAN, 2020).

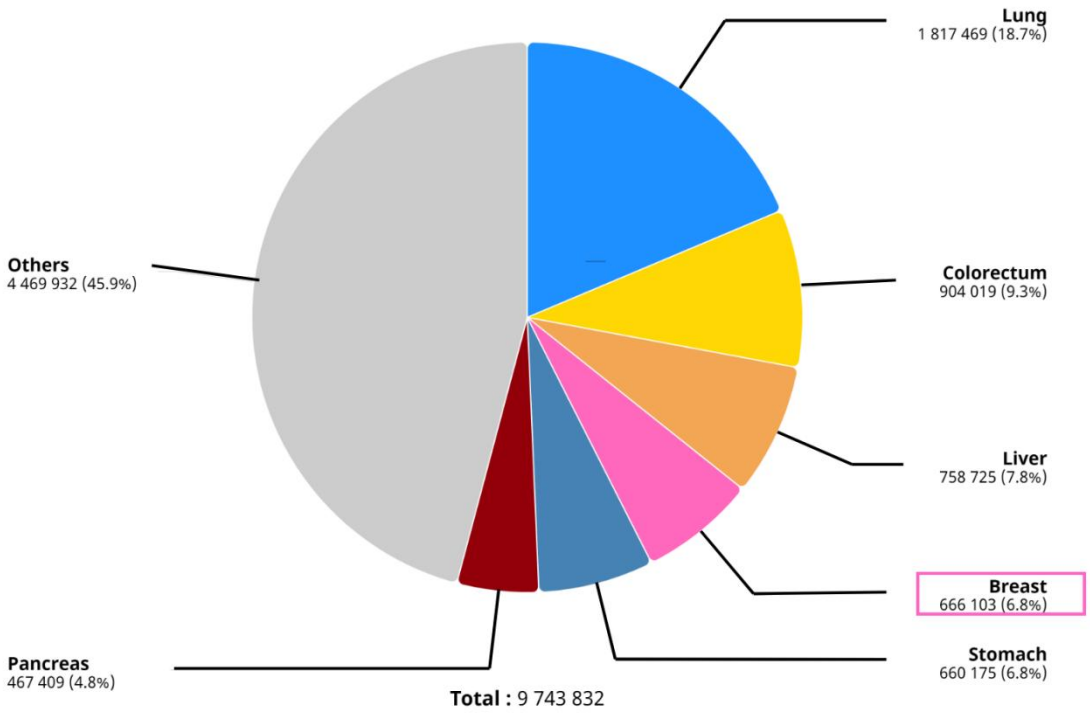

Figure 1.1: Globally estimated number of deaths due to cancer in 2022 (*GLOBOCAN*, 2022)

Breast cancer is a heterogenous health condition in terms of its origin, alterations, prognosis, and invasiveness (Ossovskaya, 2016). Breast cancer mortality is caused largely by cancer cells infiltrating adjacent tissues and metastasizing to distant organs such as the lung, bone, liver, or brain (Yang et al., 2014). Regular epithelial cells must overcome intracellular adhesion and move to distant locations in the body to spread the tumor and this action is achieved via the EMT process. Epithelial-to-mesenchymal transition is a multistep process during embryonic development (neural tube) and tissue regeneration, during which epithelial cells are transformed into mesenchymal cells. EMT is a key step in tumor invasion and metastasis, which are the leading causes behind mortality in cancer patients (Shao et al., 2015). Breast cancer stem-like cells (BCSCs) that have been produced by EMT process and are part of the primary tumor are the causative factors behind breast metastasis, cytotoxic therapy resistance and recurrence of cancer (BeLow & Clodia Osipo, 2020).

1.2 Epidemiology

As of 2020, GLOBOCAN has reported breast cancer cases to have exceeded lung cancer in women to become the most diagnosed cancer type. With 2.3 million new cases, breast cancer makes up 11.7% of cancer cases reported across 185 countries. 684,996 new cases of death from breast cancer have been recorded. It therefore ranks first for incidence in most countries (159 of 185 countries) and for mortality in 110 countries. New cases of breast cancer are more common in transitioning countries in Asia, Africa, and South America (GLOBOCAN, 2020).

From 2018 to 2022, breast cancer accounted for 15.4% of the 53.5 million prevalent cancer cases estimated globally (*Figure 1.2*). Within the same time span, more than 345,000 cancer cases were reported in Bangladesh, with 10.2% of them being breast cancer cases (*Cancer Today*, 2022).

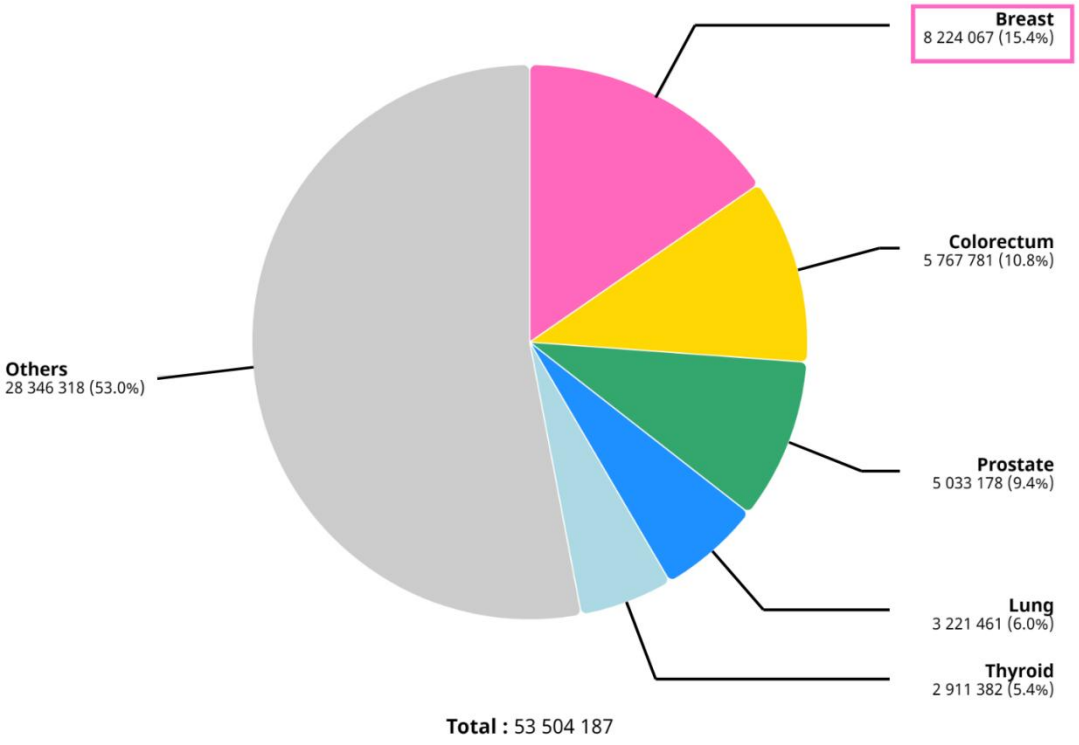

Figure 1.2: Globally estimated number of cancer cases between 2018-2022 (GLOBOCAN, 2022)

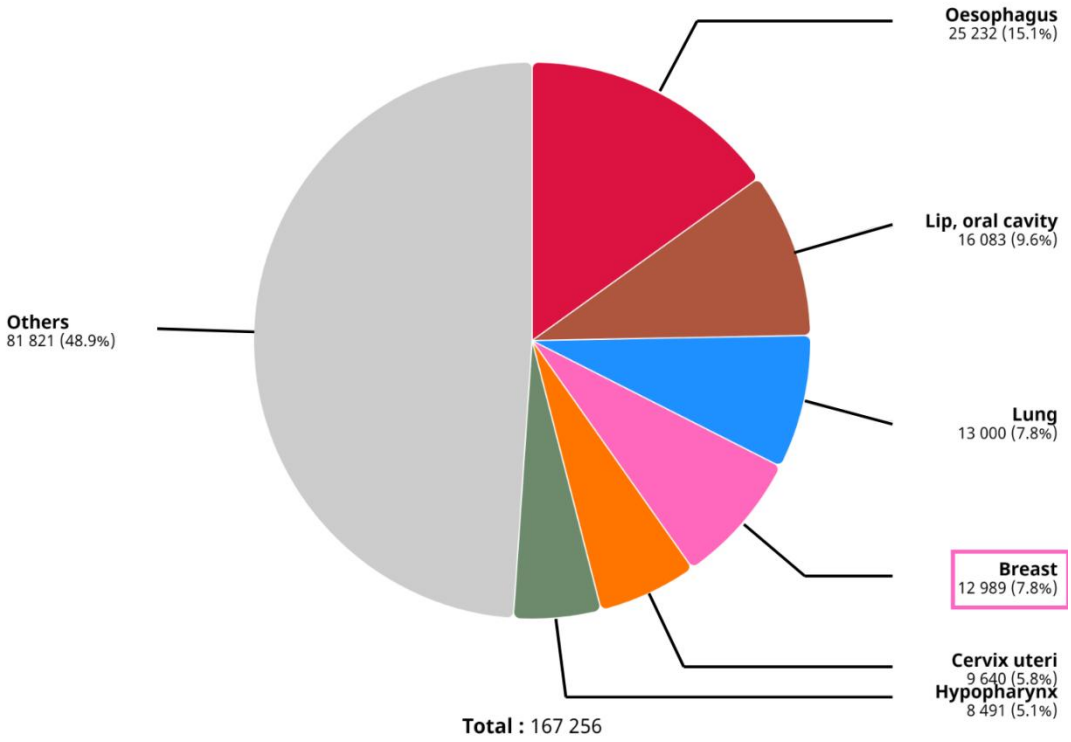

Figure 1.3: Estimated number of new cases of cancer in Bangladesh in 2022 (GLOBOCAN, 2022)

In 2022, 167,256 new cancer cases were reported in Bangladesh, where esophageal cancer made up 15.1% of the cases, followed by oral and breast cancer at 9.6% and 7.8%, respectively (*Figure 1.3*). The mortality rate was estimated at 116,598 with esophageal cancer, lung cancer, and breast cancer accounting for 20.9%, 10.3%, and 5.3% of the deaths, respectively.

1.3 Hallmarks of Cancer

Hallmarks are alterations in cell physiology or processes that turn normal human body cells into malignant and invasive tumor cells. The success of these processes is a testament to the anticancer defense mechanisms instilled within cells (Hanahan & Weinberg, 2000).

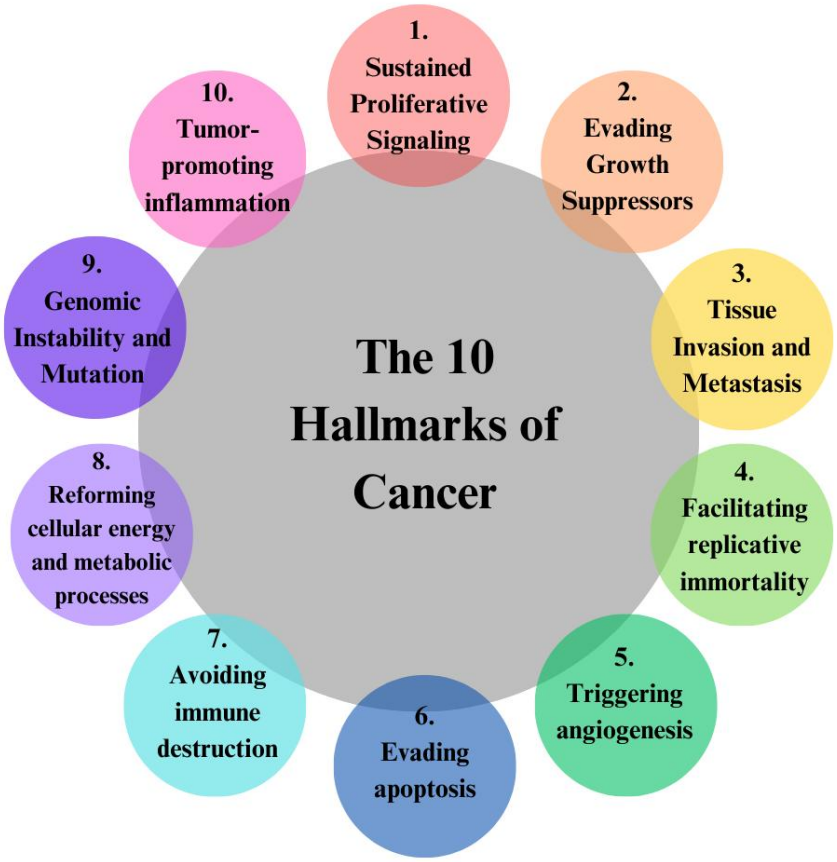

Figure 1.4: Hallmarks of Cancer

Sustained proliferative signaling: Normal cells rely on diffusible mitogenic factors such as growth signaling, in a homeostatic cell cycle to proliferate. This homeostatic cell cycle is hampered in cancer cells and results in sustained cell signaling that leads to chronic cell proliferation (Fouad & Aanei, 2017) The mutational alteration of genes within cancer cells converts them into active inducers of cell proliferation. The tumor cells eventually become capable of generating their own growth-stimulating signals, which reduces their dependence on the homeostatic mechanism and at the same time generates positive feedback signaling loop (Hanahan & Weinberg, 2000). Eventually, the tumor cells develop the ability to induce neighboring normal cells to release abundant growth stimulating signals. In certain instances, epigenetic deregulation of autocrine (auto stimulatory) and paracrine (cell-to-cell) signaling circuits can also provide cancer cells with chronic growth-promoting signals, doing so in the apparent absence of underlying somatic mutations (Hanahan & Weinberg, 2017).

Evading growth suppressors: The chronic cell division triggered by proliferative growth signals can be blocked by braking mechanisms (antigrowth signals) conducted by proteins (retinoblastoma protein (pRb)) encoded by tumor suppressor genes (TGAs) (Hanahan & Weinberg, 2017). The signals either force cells out of their proliferative stage or induce them into a post-mitotic state. The disruption of this pRb's pathway makes the cells insensitive to anti-growth factors and allows the continuation of chronic cell proliferation (Hanahan & Weinberg, 2000).

Tissue invasion and metastasis: As cancer cells become invasive and migratory, they invade adjacent cells as well as blood and lymphatic vessels, where the cells undergo metastatic growth and eventually form colonies (Hanahan & Weinberg, 2017). The migration and colonization of the cancerous cells into new locations where they have more

nutrients and space are responsible for the spreading of cancerous tumors throughout the human body (Hanahan & Weinberg, 2000).

Facilitating replicative immortality: The telomeres at the ends of chromosomes act as a third intrinsic barrier to chronic proliferation of cells. After successive cell generations, as the number of telomere repeats erodes and shortens below a certain threshold, cell cycle arrest or receptor mediated apoptosis is triggered. Cancer cells can overcome this proliferative barrier by activating a system for telomere maintenance and extension that helps to preserve the replicative capacity of stem cells. Hence, the cancerous cell becomes immortal, i.e., it achieves unlimited replicative potential for expanding its population (Hanahan & Weinberg, 2017).

Triggering angiogenesis: Angiogenesis is the formation of new blood vessels. Tumor cells need a constant supply of oxygen, glucose, and other nutrients to continue cell function and proliferation. Subsequently, the growth of developing cancer cells may stop if they are unable to acquire the nutrients from blood or when the nearest blood vessel is more than 200 micrometers away (Hanahan & Weinberg, 2017). Tumor cells activate their angiogenic switch, which they acquired during development, by altering the balance between angiogenesis inducers and its inhibitors. Altered gene transcription can carry out this activity, where the loss of p53 function causes thrombospondin-1 level to fall, thus freeing the endothelial cells from the inhibitory effects of thrombospondin-1 (Hanahan & Weinberg, 2000). Activation of the *Ras* oncogene causes the upregulation of VEGF expression.

Evading apoptosis (Resisting cell death): The tumor cell population expands in number due to continued cell proliferation accompanied by evading cell death (apoptosis). Apoptosis is propagated by sensors that monitor extracellular and intracellular environments for abnormality triggers and the effectors (caspases that are activated by death receptors) that carry out the death program by selective destruction of cellular contents and the genome. Cancer cells can evade apoptosis by mutation of p53 protein in 50% of human cancer that disarms the DNA damage sensor, hence preventing induction of effector cascade (Hanahan & Weinberg, 2000).

Avoiding immune destruction: It is expected of the human immune system to act as an efficient barrier against foreign bodies that may have entered the body. Hence, it is also responsible for the detection, neutralization and eradication of nascent cancer cells and tumors (Hanahan & Weinberg, 2011). Statistics show that 80% of human cancer cells that are non-viral can successfully circumvent immune response. These continued cases of incipient cancer cells going unnoticed by the immune systems of non-immunosuppressed individuals, raises queries regarding the reason behind the cancerous cells being able to dodge bodies' active immune surveillance. One potential explanation for the invasive nature of cancer cells is that it shares the same antigen as its healthy origin-tissue cell, which allows it to cross the immune barrier undetected. Another aspect supporting the theory elaborates that some tumor cells may eventually face recognition and attack by immune cells during later stages of cancer development. This helps them adopt immune-evasive characteristics to pass down to future cells which will be able to develop and contribute to tumor progression entirely undetected (Hanahan & Weinberg, 2017).

Reforming cellular energy and metabolic processes: Aerobic glycolysis enables cells to uptake an increased level of glucose which is subsequently metabolized by glycolysis and hence helps cancer cells meet their metabolic needs. It facilitates the formation of amino acids, nucleotides and macromolecules for proliferating cells (Liberti & Locasale, 2016). Furthermore, the capacity of cancerous cells to perform aerobic glycolysis can be acknowledged as a phenotype expressed by oncogenes as this trait is primarily caused by genes and proteins implicated in other hallmarks of cancer (Hanahan & Weinberg, 2011).

Genome Instability and Mutation: The genome is continually undermined by chemical and environmental stimuli, resulting in DNA damage that, if kept neglected, can lead to mutations and genomic instability, which further contributes to the other hallmarks of cancer (Hanahan & Weinberg, 2017). Tumorigenesis acquires growth advantages in the local tissue environment via altering oncogenes and tumor suppressor genes, and this leads to clonal development of the best-adapted cells. Increased incidences of mutation enables cells to adapt and bypass the constant monitoring by the immune system and aid in preventing apoptosis (Zhong et al., 2020). Cancer development occurs when cancer cells become more mutable due to increased susceptibility to oncogenic substances, inactivation of tumor suppressor genes, a deficiency in the genomic maintenance system, or even an impairment in immune system efficacy. The loss of telomeric DNA, which allows cancer cells to proliferate uncontrollably, is a significant contributor to genomic instability in tumors (Hanahan & Weinberg, 2011).

Tumor-promoting inflammation: Inflammatory tissue microenvironments can promote the initiation and advancement of cancer cells, leading to their acquisition of hallmark characteristics. Tumors that trigger immune response at later stages of cancer progression, are

often pervaded by infiltrating immune cells (ILC) that eventually enhance tumor advancement than eliminate it. Evidence has shown that ILCs activation can promote cancer cell proliferation, survival, evasion of anti-growth regulators, and angiogenesis. It even inhibits the action of cancer-combating cytotoxic T lymphocytes. Presence of ILCs near tumor margins can lead to increased invasiveness of tumor cells during later phases of development. Immune cells, such macrophages, have been proven to contribute to mutations and evolution of the cancer genome by exposing cancer cells to DNA-damaging reactive oxygen species (Hanahan & Weinberg, 2017).

1.4 Etiology and Risk Factors

A complicated and extensive network of controllable and uncontrollable parameters are considered to be the causative factors behind breast cancer. The complex interaction between these risk factors is a common indicator in breast cancer patients.

Age: In 2016, 99.3% of deaths due to breast cancer were among women over the age of 40 (Yi Sheng Sun et al., 2017).

Family history: Women whose mother, sister, or first-degree relatives had breast cancer are more susceptible to breast cancer due to inheritance of mutable breast cancer genes BRCA1 and BRCA2 (Yi Sheng Sun et al., 2017). Approximately 3% of hereditary breast cancers and 10% of ovarian cancers are caused by these gene mutations (Winters et al., 2017).

Breastfeeding: In a 2002 study, for each additional 12 months of breastfeeding, the risk of pre-menopausal breast cancer was reduced by 4% (Rojas & Stuckey, 2016).

Reproductive factors:

1. Early menarche (1-year delay decreases risk by 5%), late menopause (1-year delay increases risk by 3%), late pregnancy age, and low parity (number of births—each birth decreases risk by 10%) increase the risk of breast cancer (Yi Sheng Sun et al., 2017).
2. The increasing number of full-term pregnancies does decrease the overall risk of breast cancer. The risk of premenopausal breast cancer is increased by 5% and postmenopausal breast cancer by 3% for each additional year that the first full-term pregnancy is delayed. Pregnant women are at high risk of developing breast cancer due to frequent alterations in hormone levels (Rojas & Stuckey, 2016)

Estrogen Hormone: Increased exposure to the hormone increases the risk of breast cancer. Estrogen initiates the carcinogenic pathway by signaling the estrogen receptor to alter gene expression (leading to mutations) or by oxidative metabolism of estrogen (oxidative damage to DNA) (Rojas & Stuckey, 2016).

Both endogenous (produced by the ovary before menopause) and exogenous (hormone replacement therapy, oral contraceptives, if used for >10 years), estrogen increases the risk of breast cancer. However, the risk of breast cancer has been shown to significantly decrease after two years of stopping HRT (Yi Sheng Sun et al., 2017).

Lifestyle:

1. Alcohol consumption: It elevates the level of estrogen-related hormones in the blood. Intake of 35–44 grams of alcohol per day can increase the risk of breast cancer by 32% (Yi Sheng Sun et al., 2017). Breast cancer risk increases by 7% with each additional 10 g per day of alcohol consumed. Acetaldehyde, benzene, and N-nitroso dimethylamine are several carcinogens either found in alcoholic beverages or produced by alcohol

metabolism. Alcohol can also alter hormone levels by increasing circulating estrogen metabolites through suppression of hepatic estrogen metabolism and by enhancing the conversion of androgens to estrogens. Alcohol can also suppress immune function, increase cell proliferation, inhibit DNA repair, and promote cell invasion and migration (Rojas & Stuckey, 2016).

2. Dietary Fat: Overconsumption of fat (saturated fat) hampers the diagnosis and tracking of tumors in breast cancer patients (Yi Sheng Sun et al., 2017). High protein intake (red meat) may increase the risk of breast cancer by increasing the amount of circulating insulin-like growth factor-1.[4] Overweight and obese premenopausal women (40 and 49 years old) had a 14% and 26%, respectively, lower risk of developing breast cancer compared to premenopausal women who were normal weight (Winters et al., 2017).
3. Smoking: People who start smoking at an early age have greater chances of developing breast cancer (Rojas & Stuckey, 2016).

Exercise: Women who regularly engaged in strenuous physical activity (reduction in body fat leading to decreased peripheral conversion of androgens to estrogens through the enzyme aromatase) at the age of 35 years had a 14% decrease in the risk of breast cancer. In addition, the decrease in risk tends to be greater with more hours of exercise. Further, physical activity may lead to increased sex hormone–binding globulin, reducing the total amount of free circulating estrogen (Rojas & Stuckey, 2016). Fruit and vegetable consumption, as measured by plasma carotenoid levels, was inversely related to breast cancer risk, especially among non-obese women (Winters et al., 2017).

Radiation: Delayed adverse effects of chest radiation received at a young age can contribute to tumor formation and invasion (Rojas & Stuckey, 2016).

1.5 Molecular Subtypes of Breast Cancer

The molecular classification of breast carcinoma has been divided into 5 subtypes:

1. Luminal-A subtype breast cancer
2. Luminal-B subtype breast cancer
3. HER2-positive breast cancer
4. TNBC (triple negative breast cancer)

Luminal A type breast cancer accounts for 50% of invasive breast cancer cases, whereas **Luminal B** type breast cancer makes up only 20% of all cases. Both exhibit positive expression of ER/PR (estrogen receptor/progesterone receptor) and negative expression of HER2 (human epidermal growth factor receptor 2) and are significantly responsive to endocrine therapy. Although the Luminal-A type has the best prognosis of all subtypes, both can be treated by targeting protein markers.

HER2-positive breast cancer is associated with high expression of HER2 gene and negative expression of ER/PR. Despite its poor prognosis, it can be treated by targeting protein markers. It is usually responsive to treatment involving trastuzumab.

TNBC is referred to as a subtype of basal-like breast cancer and accounts for 15% of all invasive breast cancers. It has negative expressions for ER, PR, and HER2 and as a result, offers no response to endocrine therapy or trastuzumab. Due to TNBC having the worst prognosis, patients have a shorter survival time (mortality rate is 40% within 5 years of diagnosis). The mortality rate of TNBC patients within 3 months after recurrence is as high as 75%. As it cannot be treated by targeting protein markers, chemotherapy is the only line of therapy. In instances of highly invasive cases, patient survival can be as low as 13 months

post-metastasis. As it cannot be treated by targeting protein markers, chemotherapy is the only line of therapy. Triple-negative breast cancer is a kind of breast cancer that does not have any of the receptors that are commonly found in breast cancer; hence, treatment options are limited (lumpectomy, mastectomy, chemotherapy, radiation) (CDCBreastCancer, 2023). The incidence of TNBC has often been consistent with overexpression of cytokeratin CK5/6 and CK14/17, cyclin-D1, and P-cadherin, caveolin 1 and 2, as well as p53 alterations. Data from studies have indicated that the NF-κB pathway and its components may be a key regulator of TNBC. This pathway is responsible for maintenance of immune response, extracellular matrix degradation, angiogenesis, the cell cycle and apoptosis (Ossovskaya, 2016). Additionally, activation of the PI3K/Akt/mTOR pathway has a 10-21% effect on cell cycle control, growth, and inactivation in TNBC (Citra Dewi et al., 2022).

Table 1.1: Molecular Classification of Breast Cancer (Orrantia-Borunda et al., 2022)

Molecular Sub-type	**Luminal A**	**Luminal B**	**HER2**	**TNBC**
Prevalence among breast cancer cases	50%	20%	15%	15%
Tumor growth rate	Slow	Faster than Luminal A	Fast	Fast and aggressive
Prognosis	Good	Intermediate	Poor	Poor
Targeted therapy	Hormonal (endocrine) therapy and targeted therapy	Hormonal (endocrine) therapy and targeted therapy	HER2-targeted therapy and usually responsive to treatment using trastuzumab	Chemotherapy
Receptor/Protein expression	Positive expression of ER/PR and negative expression of HER2	Positive expression of ER/PR and negative expression of HER2	High expression of HER2 gene and negative expression of ER/PR	Negative expressions of ER, PR, and HER2
Cellular marker expression	Low expression of Ki-67	High expression of Ki-67	Low expression of Ki-67	Low expression of Ki-67

1.6 NOTCH Signaling Pathway

The NOTCH pathway is a highly conserved cell signaling pathway conducted through a single-pass type-1 transmembrane receptor protein. It regulates cell proliferation, tumor suppression, cell fate, differentiation, and apoptosis. Hence, erroneous NOTCH signaling can result in overexpression or abnormal activity of its receptors (Panagiotis Ntziachristos et al., 2014). Studies have established that Notch signaling contributes to drug resistance and enhances BCSC survival (BeLow & Clodia Osipo, 2020). NOTCH proteins are transmembrane receptors that are often triggered by transmembrane ligands generated on adjacent cells. NOTCH receptors in their mature state are heterodimers composed of a transmembrane component and an extracellular subunit (Majumder et al., 2020). Ideally, it spans the cell membrane once, with its N-terminus on the external side of the membrane (*UniProt*, 2024). The NOTCH pathway works by communication between two cells, where one is the signal receiving cell (more NOTCH receptor molecules than ligand molecules) and the other is the signal sending cell (more ligand molecules than Notch receptor molecules). This pathway is activated by direct contact between cells, where the transmembrane proteins of a cell produce and secrete ligands to extracellular regions that bind to the extracellular NOTCH domain of its adjacent cell. This binding trigger proteolytic events that cause the intracellular domain to detach from the protein and subsequently enter the nucleus to control gene expression. The 5 NOTCH ligands are: Delta-like 1, Delta-like 3, Delta-like 4, Jagged 1, and Jagged 2 (Panagiotis Ntziachristos et al., 2014). Transmembrane proteins (with EGF-like proteins) such as delta-like and jagged usually trigger the notch signaling pathway.

The signal sending cell contains the Delta/jagged ligand, which (otherwise inactive) must be activated so that it can interact with the NECD domain of the NOTCH receptor. There are 4 NOTCH receptors: NOTCH1, NOTCH2, NOTCH3, and NOTCH4 that are constructed of 3

functional components: N-terminal extracellular (NECD), C-terminal transmembrane-intracellular subunit (NTM), and intracellular (NICD) domains. The NECD lies on the external part of the cell, and the NICD lies in the internal part of the cell (cytoplasm), while these two domains are interconnected by the NTM that lies on the cell membrane. The protein extends across the cell membrane, thus enabling both extracellular and intracellular functions of the protein. In the presence of the E3 ubiquitin ligase enzyme, the Mib (Mind Bomb) protein triggers ubiquitination, where the ubiquitin protein binds to the delta ligand of the signal sending cell to activate it. The activated ligand then repositions itself so that it can bind to the NECD domain of the signal receiving cell. This interaction between ligand and NECD causes the cell to release TACE (TNF-α ADAM metalloprotease converting enzyme) that cleaves NECD from the reception site by a process called S2 cleavage. This leaves only the NTM domain and NICD of the Notch receptor at the membrane site (Panagiotis Ntziachristos et al., 2014). The Ligand-receptor binding creates a physical force that pulls on the NOTCH receptor and exposes the S2 site to proteolytic cleavage by an ADAM metalloprotease (mostly *ADAM10*). S2 cleavage then generates a transmembrane intermediate that becomes sensitive to proteolysis by the γ-secretase complex at a third intramembrane site (S3) (Wolfe, 2020). A ɣ-secretase enzyme carries out S3 cleavage by cleaving off NICD from the NTM domain. The resulting free NECD in the cytosol forms a complex with the DNA binding factor CSL (ubiquitous DNA-binding protein), or Mastermind, and the P300 (transcription factor) protein. The complex translocates to the nucleus, where P300 acts as histone acetylase in the transcription of NOTCH target genes (Panagiotis Ntziachristos et al., 2014). The type of target gene activated by this pathway depends on the cell-type and ligand-receptor interaction between signal-sending and receiving cells. Some common target genes activated by NOTCH signaling are *cyclinD1, MYC, p21, BCL2* (Previs et al., 2015).

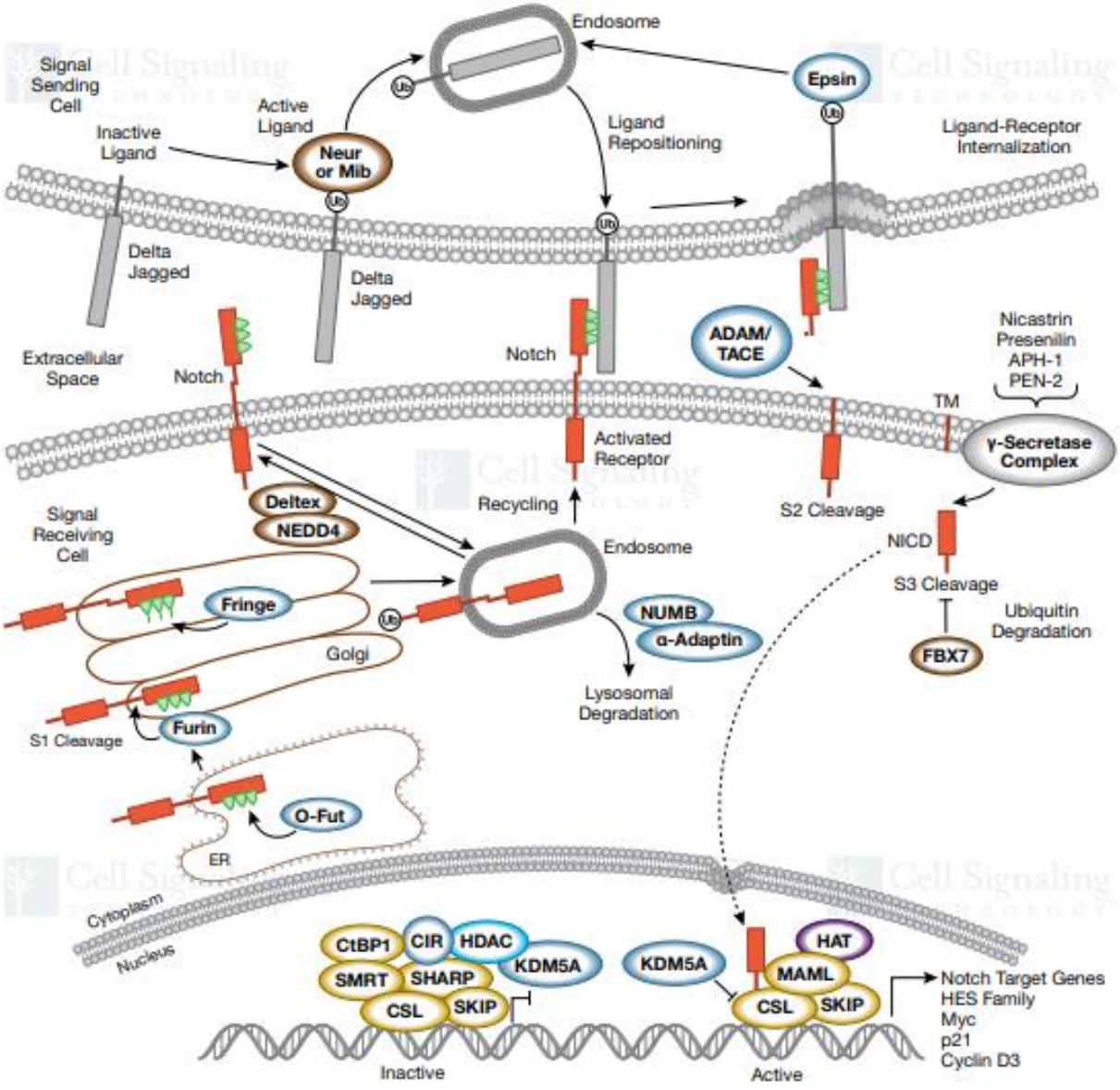

Figure 1.5: NOTCH Signaling Pathway (*Cell Signaling Technology*, 2014)

NOTCH pathway's oncogenic effects include apoptosis inhibition and drug resistance, induction of transition (EMT), epithelial-mesenchymal maintenance of a stem-like phenotype, and metastasis, in addition to induction of angiogenesis and other tumor-stroma interactions that are less specific to individual tumor types. There is also evidence that NOTCH can act as an inhibitor of tumors in some cases, suggesting that the mechanisms involved are complicated and context sensitive. It also mediates Notch signaling, inhibits or promotes VEGF-induced endothelial cell proliferation during angiogenesis (Majumder et al., 2020).

The upregulation of NOTCH signaling pathway activity provided the basis of its (Viatour et al., 2011). During normal somatic cell development, the differentiation of breast epithelial cells is regulated by the NOTCH pathway. In regard to that, this particular pathway of interest has been proven to have both tumor suppressive and oncogenic effects, depending on the expression patterns, state of the associated proteins and trigger factors in variable microenvironments. Hence, different alterations in Notch pathway can suppress or promote different types of breast cancer (Nandi & Chakrabarti, 2020). Studies have determined that NOTCH1, NOTCH4 receptors and JAG1 (ligand) impose oncogenic effects whereas NOTCH2 exhibits both oncogenic and tumor suppressive effects upon breast cancer (Previs et al., 2015).

1.7 Notch regulated EMT in Breast cancer

NOTCH signaling strongly regulates the activity of its downstream gene NF-κB (nuclear factor kappa B), while NF-κB has been identified to play an important role in the EMT process. As a precondition for these cells to undergo EMT toward an invasive, metastatic tumor phenotype, NF-B protects mammary epithelial cells against TGF-induced cell death. Its inhibition makes the cancer cells sensitive to anticancer drug-induced apoptosis (Huber et al., 2004). The NOTCH-DELTA signaling pathway triggers the upregulation of transcription factors (genes) such as *Zeb1, TWIST, SNAIL*, which are some of the many trigger factors that activate the transition process. *Snail, Slug, and ZEB2/SIP1* are direct transcriptional repressors of E-cadherin, whereas *Twist and ZEB1* act indirectly on E-cadherin (Polyak & Weinberg, 2009). These transcription factors modulate the expression of several proteins that are involved in cell polarity, cell-to-cell interaction, the cytoskeleton framework, and extracellular matrix breakdown (Yang et al., 2014). It also increases the production of matrix

metalloproteinases that dissolve the extracellular matrix and allow the cells to detach from the basal membrane and eventually migrate.

Inhibition of NOTCH1 reverses the EMT process both *in vitro* and *in vivo and* inhibits migration and invasion in breast cancer cells to distant regions. Overexpression of NOTCH1 and/or Jagged1 indicates a poor prognosis for breast cancer patients. Blocking NOTCH1 could reverse the EMT pathway and lead to MET. The MET program can be detected by increased E-cadherin and occludin expression and reduced N-cadherin, vimentin, and nuclear β-catenin expression, thus resulting in the development of an epithelial phenotype. Meanwhile, NOTCH1 inhibition can lower breast cancer cell migration and invasion. Likewise, Jagged1-induced NOTCH signaling activation has been identified as a major driver of the EMT process. This EMT program resulted in cancer cells changing into a highly invasive and mesenchymal phenotype by drastically reducing E-cadherin and occludin expression and significantly increasing N-cadherin, vimentin, and nuclear β-catenin expression. According to a recent study, EMT induced E-cadherin loss subsequently compromises cell-cell adhesion and permits β-catenin nuclear localization (Shao et al., 2015).

1.8 Bioinformatic Analysis of NOTCH Pathway in Breast Cancer

For several decades, the NOTCH signaling pathway has been an area of interest for drug developers. Various drugs targeting a different component of this complex network have been designed and tested in preclinical and clinical settings, but none has been approved as of yet. Current studies have focused upon the effect of NOTCH signaling on tumorigenesis and cell survival and with respect to that, this study was planned to shed light on the specific genes that are correlated with this pathway, their differential expression in breast cancer and

lastly, their ability to suppress cancerous tumor progression when regulated via external chemical entities, such as drugs.

The concept of bioinformatics combines the field of biology and information technology and hence, bioinformatic analysis is the utilization of analytical and computational tools to collect and interpret biological data. It can be conveniently listed as a method of primary data collection in the field of biology and medicine. This technique can be used to analyze genome sequence, gene variation and expression, predict gene and its respective protein's structure as well as representation and evaluation of biological pathways for better comprehension of gene-disease interactions (Bayat, 2002).

1.9 Role of SNPs in Breast Cancer

Single nucleotide polymorphisms (SNPs) are sequence variations in the human genome stemming from point mutations within a DNA sequence. As the mutation alters, deletes, or inserts a single nucleotide base from the sequence, an allele of the DNA sequence is formed. SNPs interfere with protein function that usually results in formation of an unstable or faulty protein. A SNP is defined as having a minor allele frequency of more than 1% in at least one group (Erichsen & Chanock, 2004). The polymorphism may take place in either coding or non-coding regions of a gene.

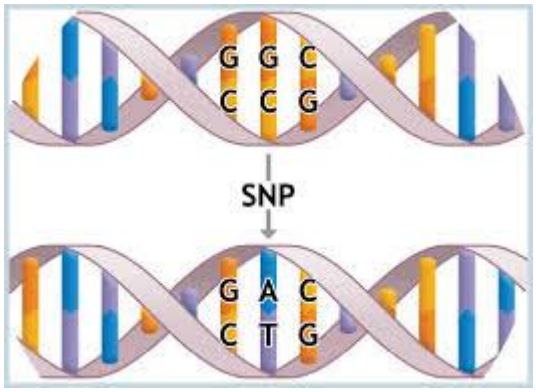

Figure 1.6: Single Nucleotide Polymorphism (SNP) (*Society for Mucosal Immunology*, 2024)

The occurrence of SNPs in non-coding regions can modify the structure of mRNA involved in protein synthesis or even change the expression level of the gene. SNPs are one of the main factors behind genetic variation in the human genome, hence, it has been reported to increase the risk of cancer (Zhu et al., 2023).

SNPs are of particular importance in the fields of medicine and bioinformatics due to their role as therapeutic and clinical biological markers. As the polymorphism occurs in DNA between genes, it can help researchers identify the genes that are related to the disease being studied. In this study, the determination of probable functional SNPs of a promising gene will help analyze the gene’s response to variable toxins, environmental factors and tumor susceptibility. SNPs may indicate a person's response to specific drugs or chemicals and likelihood of acquiring illnesses (Shastry, 2007).

Chapter 2

Aims and Objectives

The NOTCH signaling pathway in breast cancer is not an exclusive pathway with genes and proteins distinct from other cellular pathways. There are far too many genes that have been observed to be common between multiple pathways. These genes enable the activation of one pathway to directly or indirectly influence the initiation, alterations or outcome of another pathway they are also a part of. For instance, the activation of VEGFR2 (protein kinase) by the VEGF pathway triggers the activation of PLCγ followed by PKC, which eventually stimulates the initiation of MAPK signaling pathway. As a result, nuclear DNA synthesis is stimulated, and subsequent proliferation of endothelial cells takes place. Similarly, it can be assumed that the genes or ligands switched on by NOTCH pathway are likely to influence other pathways into bringing about a cellular alteration.

This study is designed to identify a few genes that fall under two specifications:

1. They are likely to have a significant effect on other signal transduction processes.
2. Genes acting on other signal transduction processes can lead to outcomes that affect breast cancer tumor proliferation.

Further analysis will focus on the functional SNPs of the most promising genes of breast cancer. Diseases such as breast cancer possess a genetic aspect to it, i.e. presence of a genetic component has the ability to increase or decrease the body's susceptibility to the disease. Accurate identification of the genes will carve a path for further study into development of drugs that will explicitly target those genes or ligands to exert anti-tumor activity in people diagnosed with breast cancer.

Chapter 3

Methodology

3.1 Data Sources

The outcome of this thesis was constructed on the basis of primary data collected from KEGG Database, GSEA, UALCAN, cBioPortal for Cancer Genomics, GEPIA 2, Ensembl Genome Browser 111, ShinyGo v0.741 and SNPs3D .

KEGG Database (Kyoto Encyclopedia of Genes and Genomes)**:** It is a database that uses information for credible sites and published literature to establish manually created pathway maps. Thus, it is a source representing the various molecular networks of our biological system. Information about genes, proteins, genomes, biochemical reactions, enzyme nomenclature, drugs and diseases are also accessible through KEGG Database. It was used to obtain the names of different signal transduction pathways and systems associated with the human body. In addition to that, the genes involved in each of those pathways were collected. (https://www.genome.jp/kegg/)

GSEA (Gene Set Enrichment Analysis): This tool is used to carry out enrichment analysis of two groups of a gene set that had been subjected to variable biological states. This helps to determine whether the differences between the two groups are significant or by chance. (https://www.gsea-msigdb.org/gsea/index.jsp)

UALCAN (The University of Alabama at Birmingham Cancer data analysis Portal): This website resources cancer OMICS data on different cancer types, its related genes and their expression profile, survival analysis, correlation etc. The site also provides users access to data represented in the form of graphs and plots.

(https://ualcan.path.uab.edu/)

cBioPortal for Cancer Genomics: It is an analytical tool that gives users access to large volumes of genomic data on cancer. Its primary source of information is TCGA (The Cancer Genome Atlas), TARGET, and published papers and articles from reliable sources. This website also contains other gene-related data such as correlated genes, survival profile, possible profile, its related pathways etc.
(https://www.cbioportal.org/)

GEPIA 2 (Gene Expression Profiling Interactive Analysis): A gene analysis tool that has data related to its expression in different cancer types. Additionally, it also comprises information on differentially expressed genes in a particular type of cancer, along with their Log2FC and adjusted p-value.
(http://gepia2.cancer-pku.cn/#index)

Ensembl Genome Browser 111: This database contains information related to gene variation, genomics and transcriptional regulation. It was used to determine the Ensembl ID of query genes that was later used for gene set enrichment analysis
(http://www.ensembl.org/info/genome/stable_ids/index.html)

ShinyGo v0.741: It is a gene-set enrichment tool that represents data in graphical format. It collects and interprets information from various other sites such as KEGG Database, Ensembl etc to display pathway diagrams, genome plot, protein-protein interaction, promoter motif and gene characteristics of the input gene set.
(http://bioinformatics.sdstate.edu/go74/)

SNPs3D: A database dedicated to information about SNPs (single nucleotide polymorphisms) of a gene, their SNP id, svm profile, molecular effect and other related information to determine the possible genetic variations of a gene.

(http://www.snps3d.org/)

3.2 Literature Search

A thorough systematic literature assessment was conducted to choose appropriate primary information from peer-reviewed research papers available on PubMed, Springer, ScienceDirect, Google Scholar, NCBI, Elsevier, ResearchGate etc. sites and journals. The database of each of these sites were searched using keywords such as 'Breast cancer classification and clinical study', 'Anti-tumor activity', 'NOTCH signaling pathway', 'Genes associated with NOTCH pathway in breast cancer', 'Cell signaling pathways', 'Correlation between NOTCH signaling and other cell signaling pathways'. The obtained data was then filtered to choose only the most relevant information, followed by arranging all collected data according to the paper's outline. The study emphasized using authentic and reliable information and accordingly, all relevant literary works were property cited.

3.3 Data Collection Method

This paper aimed to analyze the different genes and proteins that are part of the NOTCH signaling pathway, the extent to which they affect other cell signaling pathways and its possible effect on breast cancer tumor formation, prognosis, growth, control and inactivation. Similarly, it is imperative to identify, compare and study the genes that are differentially expressed in breast cancer and are correlated with prominent NOTCH genes: *NOTCH1, NOTCH2, NOTCH3, and NOTCH4.*

The primary data was collected in the following order:

1. **L1**- List of correlated genes (both positively and negatively correlated) with *NOTCH1, NOTCH2, NOTCH3, and NOTCH4* respectively in breast cancer. The genes were recorded and filtered for further processing if they fell within the Pearson Correlation Coefficient threshold of $-0.3 \geq x \geq 0.3$. The information for 4,023 correlated genes was collected from UALCAN database.
2. **L2**- List of pathways and systems that are associated with signal transduction in the body and the list of genes involved in each pathway were extracted from KEGG Database. A total of 19 pathways were listed.
3. **L3**- List of co-expressed genes with *NOTCH1, NOTCH2, NOTCH3, and NOTCH4* in breast cancer. A total of 20,055 genes were recorded and filtered for further processing if they fell within the Spearman Correlation Coefficient threshold of $-0.4 \geq x \geq 0.4$ and the p-value was lower than 0.05. In addition to that, the specific Spearman CC and p-value for each gene was also recorded. The analytical tool cBioPortal was the source of this data.
4. **L4**- List of differentially expressed genes (both over-expressed and under-expressed) in breast cancer. The list was filtered on the basis of Log2FC cutoff of 1 and a q-value cutoff of 0.05. The Log2FC value and p-adjusted value for each gene was recorded. This information for 3560 genes was collected from GEPIA 2 database.

The initial data was collected from UALCAN to identify the genes that are positively and negatively correlated with *NOTCH1, NOTCH2, NOTCH3 and NOTCH4* in breast cancer. This data (L1) was cross-checked with the gene sets of other cell-signaling pathways that were collected from KEGG Database (L2). This was done to detect and record the genes that are common between NOTCH signaling and other signaling pathways into a new list, L5

(Table 4.1). For the new list of common genes (L5) between L1 and L2, their associated pathways and Pearson CC were subsequently recorded.

Another set of data was collected from cBioPortal consisting of the genes that are co-expressed with *NOTCH1, NOTCH2, NOTCH3, and NOTCH4* in breast cancer and this list was labeled L3. Each gene listed previously in L5 was then individually cross-checked with L3. The new list of common genes between L5 and L3 was labeled L6 and for each gene in this list, their associated Spearman Correlation Coefficient and p-value were also recorded.

The last set of data was collected from GEPIA 2 and it was a list of genes that are differentially expressed (both over-expressed and under-expressed) in breast cancer with a $|Log_2FC|$ cutoff of 1 and a q-value cutoff of 0.05. This new gene list was labeled L4. The genes in the previously derived list L6 were cross-checked with the differentially expressed gene list (L4) to form a filtered list of genes, L7. Data from GEPIA2 was used to assign the specific Log2FC value and p-adjusted value for each gene in L7.

The resulting gene list L7 *(Table 5.1)* consists of breast cancer genes that are all:

1. Correlated with either *NOTCH1, NOTCH2, NOTCH3 or NOTCH4*
2. Associated with at least one cell-signaling pathway alongside NOTCH signaling pathway.
3. Significantly correlated with NOTCH and other cell-signaling pathways.
4. Differentially expressed in breast cancer.

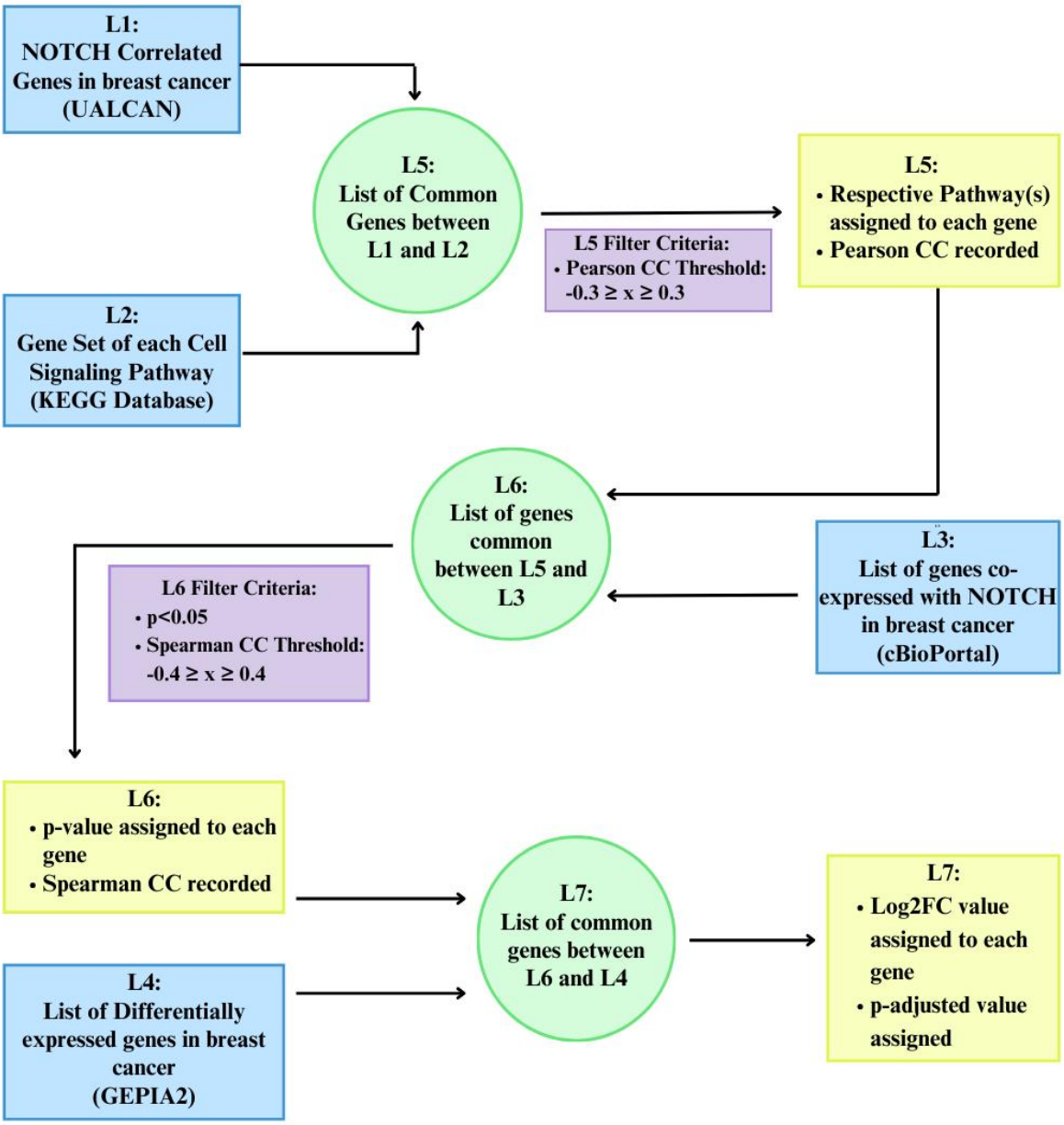

Figure 3.1: Schematic Representation of Primary Data Collection Method

Chapter 4

Primary Data

4.1 NOTCH Correlated Genes [L1]

The UALCAN website was used to extract two complete lists of genes that are positively and negatively (respectively) correlated with *NOTCH1* in breast cancer. The Pearson CC value for each gene was also noted in the list. Only genes with correlation coefficient within the threshold of $-0.3 \geq x \geq 0.3$ were included in the list. The same method was repeated for *NOTCH2, NOTCH3, and NOTCH4*.

4.2 Gene Sets of Signal Transduction Pathways [L2]

A list of 19 different signal transduction pathways (including NOTCH signaling) of human body was determined from KEGG PATHWAY Database and this website was then used to extract the complete gene set for each of these pathways. The signal transduction pathways are:

1) MAPK signaling pathway
2) ErbB signaling pathway
3) Rap1 signaling pathway
4) Wnt signaling pathway
5) Notch signaling pathway
6) TGF-beta signaling pathway`
7) Hippo signaling pathway
8) VEGF signaling pathway
9) JAK-STAT signaling pathway
10) NF-kappa B signaling pathway

11) TNF signaling pathway

12) HIF-1 signaling pathway

13) FoxO signaling pathway

14) Phosphatidylinositol signaling system

15) Sphingolipid signaling pathway

16) cAMP signaling pathway

17) PI3K-Akt signaling pathway

18) AMPK signaling pathway

19) mTOR signaling pathway

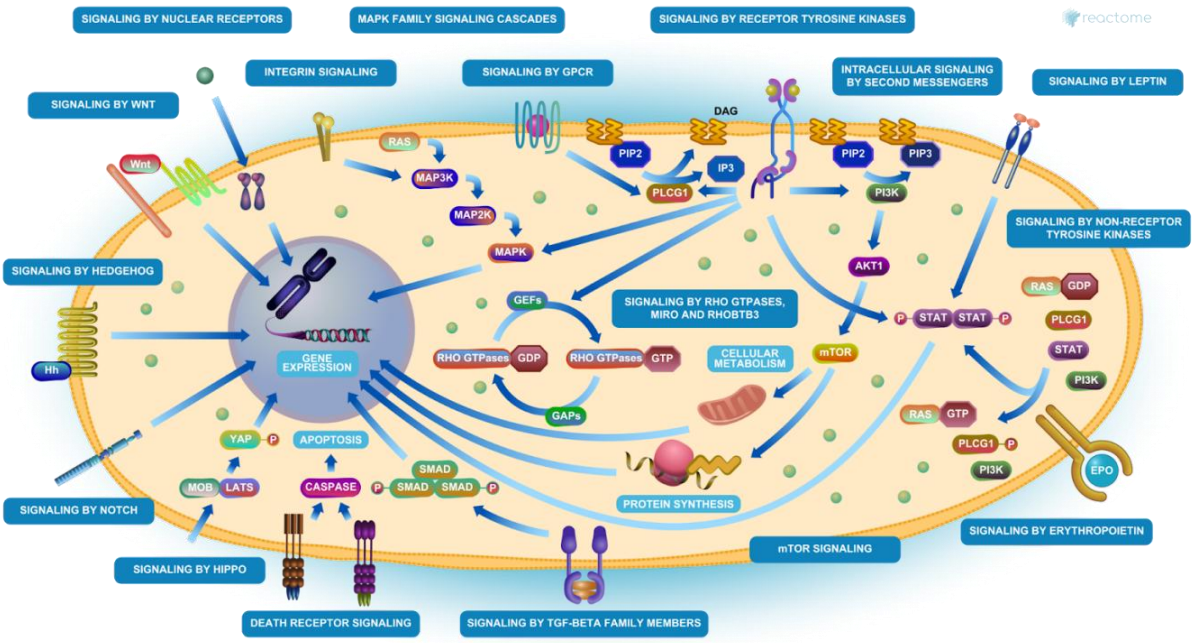

Figure 4.1: Signal Transduction Pathways (*Reactome, Signal Transduction*, 2023)

4.3 Common genes between the NOTCH Correlated Gene Set and Signal Transduction genes

Table 4.1 below accounts for genes that were found to be common between the NOTCH Correlated Gene Set and the Gene Sets of Signal Transduction Pathways. These NOTCH genes are either over-expressed or under-expressed in breast cancer and affect tumor cell proliferation by regulating receptors, ligands or genes of the other signaling pathways they have been found to be associated with. The association signifies that these genes can regulate an aspect of the pathway and eventually lead to tumor cell growth proliferation or inhibition.

Table 4.1: Common genes between the NOTCH Correlated Gene Set and Signal Transduction genes (L5)

NOTCH1

Positively Correlated Genes	Pathway	Pearson CC
DUSP7	MAPK	0.48
FSCN1	AMPK	0.46
NCK2	Erbb	0.46
PRKD3	VEGF	0.46
FLNA	MAPK	0.45
FAT1	WnT	0.43
TCF7L1	WnT	0.43
WWTR1	Hippo	0.43
PLD1	mTOR	0.41
CDH3	WnT	0.4
TLE4	WnT	0.4
PLCG2	VEGF, Erbb, Phosphatidylinositol	0.39
ADAM17	TNF	0.38
AMOTL1	Hippo	0.38

Table 4.1 (continued)

NOTCH1

Positively Correlated Genes	Pathway	Pearson CC
MYC	Erbb, HIF-1, AMPK	0.38
NFATC1	WnT, AMPK	0.38
PLK1	FoxO, AMPK	0.37
POLDIP3	mTOR	0.37
SMURF1	TGF-beta	0.37
CFLAR	AMPK	0.36
FZD7	WnT	0.36
ITPR3	WnT, Phosphatidylinositol	0.36
LPIN1	HIF-1	0.36
PIK3CD	VEGF, Erbb, Phosphatidylinositol	0.36
SKI	TGF-beta	0.36
SOX9	cAMP	0.36
TGFA	Erbb	0.36
AKT3	MAPK, Erbb	0.35
IRAK1	NF-kappa B	0.35
MAPK7	cAMP	0.35
NCK1	Erbb	0.35
PCDHB10	WnT	0.35
SFRP1	WnT	0.35
ETS1	VEGF	0.34
FOXM1	AMPK	0.34
IMPA2	Phosphatidylinositol	0.34
MAP4K4	NF-kappa B, TNF	0.34
PCDHB2	WnT	0.34
SMARCA4	HIF-1, WnT	0.34
EDN1	HIF-1	0.33
INPP5E	Phosphatidylinositol	0.33

Table 4.1 (continued)

NOTCH1

Positively Correlated Genes	Pathway	Pearson CC
NPHP4	Hippo	0.33
PCDHB14	WnT	0.33
PML	mTOR	0.33
CEBPB	AMPK	0.32
CSK	AMPK	0.32
LATS2	Hippo	0.32
TCF7L2	WnT	0.32
ADORA2B	HIF-1	0.31
BMP6	TGF-beta	0.31
DVL2	Hippo, WnT	0.31
PLD2	mTOR	0.31
SKP2	FoxO	0.31
SOS1	Erbb	0.31
SPHK1	VEGF, Sphingolipid	0.31
CBL	Erbb	0.3
CRTC2	AMPK	0.3
LRP6	WnT	0.3
MKI67	HIF-1	0.3
SLC2A1	HIF-1	0.3
TLE1	WnT	0.3
Negatively Correlated Genes	**Pathway**	**Pearson CC**
GATA3	HIF-1	-0.32

Table 4.1 (continued)

NOTCH2

Positively Correlated Genes	Pathway	Pearson CC
CBL	Erbb	0.48
RRAGC	mTOR	0.48
ELK4	MAPK	0.47
ROCK1	HIF-1	0.47
CRK	Erbb, MAPK	0.46
EP300	FoxO, TGF-beta, HIF-1, WnT, AMPK	0.46
LATS1	Hippo	0.46
RAPGEF2	cAMP	0.46
SP3	HIF-1	0.46
YAP1	Hippo, AMPK	0.46
SP1	HIF-1	0.45
MAPK14	VEGF, TGF-beta, AMPK	0.44
TJP1	Hippo	0.44
ABL2	Erbb, AMPK	0.43
APPL1	AMPK	0.43
ATR	HIF-1	0.43
CREBBP	FoxO, TGF-beta, HIF-1, WnT	0.43
PAK2	Erbb	0.43
ARID1A	WnT	0.42
ATF2	MAPK, TGF-beta	0.42
ADAM17	TNF	0.41
GNG12	WnT, MAPK	0.41
MAPK1	VEGF, mTOR, Erbb, TGF-beta, AMPK	0.41
PIK3C2A	VEGF, Phosphatidylinositol, HIF-1	0.41
ACVR1	TGF-beta	0.4

Table 4.1 (continued)

NOTCH2

Positively Correlated Genes	Pathway	Pearson CC
ERCC6	HIF-1	0.4
GBF1	AMPK	0.4
GSK3B	Erbb, AMPK, WnT	0.4
RICTOR	PI3K-Akt	0.4
SMARCA5	WnT	0.4
ATM	AMPK	0.39
NFATC3	WnT	0.39
SMAD4	TGF-beta, WnT	0.39
SMURF1	TGF-beta	0.39
ATF1	AMPK	0.38
INO80	WnT	0.38
OSBP	Sphingolipid	0.38
PRKAA1	AMPK	0.38
PRKAB2	AMPK	0.38
PRKG1	Rap1	0.38
RIPK1	TNF, NF-kappa B	0.38
STAT3	JAK-STAT, AMPK, HIF-1	0.38
TAB2	TNF	0.38
LATS2	Hippo	0.37
NFKB1	NF-kappa B, TNF, HIF-1	0.37
RNF115	AMPK	0.37
SMAD2	TGF-beta	0.37
CHUK	NF-kappa B, TNF, FoxO, MAPK	0.36
CSNK1G1	FoxO	0.36
CYLD	TNF	0.36
DGKH	Phosphatidylinositol	0.36
FNIP2	AMPK	0.36

Table 4.1 (continued)

NOTCH2

Positively Correlated Genes	Pathway	Pearson CC
GNAQ	WnT	0.36
MAPK8	FoxO, Erbb, TGF-beta, AMPK	0.36
NFE2L2	HIF-1	0.36
PTEN	AMPK, Phosphatidylinositol	0.36
SRCAP	WnT	0.36
TRAF6	NF-kappa B	0.36
AMOTL1	Hippo	0.35
ARNT	HIF-1	0.35
BIRC2	TNF	0.35
EIF4EBP2	cAMP	0.35
JAK1	JAK-STAT	0.35
NCK1	Erbb	0.35
Negatively Correlated Genes	**Pathway**	**Pearson CC**
ATPIF1	HIF-1	-0.31
HRAS	VEGF, mTOR, Erbb, TGF-beta	-0.31
MIF	AMPK	-0.31

NOTCH3

Positively Correlated Genes	Pathway	Pearson CC
BMP6	*TGF-beta*	*0.54*
KSR1	*AMPK*	*0.38*
PLD1	*mTOR*	*0.36*

Table 4.1 (continued)

NOTCH4

Positively Correlated Genes	Pathway	Pearson CC
FLT4	HIF-1	0.78
ACVRL1	TGF-beta	0.67
PDE2A	cAMP	0.67
ENG	HIF-1	0.64
PPAP2A	Sphingolipid	0.64
PCDH12	WnT	0.63
NOS3	VEGF, AMPK	0.58
RAPGEF3	Rap1, cAMP	0.58
TEK	HIF-1	0.58
AQP1	HIF-1	0.57
CAV1	TNF	0.57
TBXA2R	AMPK	0.53
EPAS1	HIF-1	0.5
EDNRB	HIF-1	0.49
KDR	VEGF, HIF-1	0.49
PGF	HIF-1	0.49
LIPE	AMPK	0.47
TGFB1	HIF-1, AMPK	0.47
GLI1	AMPK	0.45
IGF1	AMPK, HIF-1	0.45
PPAP2B	Sphingolipid	0.45
FLT1	HIF-1	0.44
PCDHGC3	WnT	0.43
ACACB	AMPK	0.42
JUNB	TGF-beta	0.42

Table 4.1 (continued)

NOTCH4

Positively Correlated Genes	Pathway	Pearson CC
MAPK11	TGF-beta	0.42
TWIST2	HIF-1	0.42
DUSP1	MAPK	0.41
FHL1	HIF-1	0.41
PNPLA2	AMPK	0.41
CXCL12	HIF-1	0.4
DCHS1	WnT	0.4
JUN	Erbb, AMPK, TGF-beta	0.4
SFRP2	WnT	0.4
ADAMTS1	HIF-1	0.39
PPARG	AMPK	0.38
SFRP4	WnT	0.38
PCDHGA6	WnT	0.37
SHC2	VEGF, Erbb	0.37
SOCS3	AMPK	0.37
TGFBR2	TGF-beta	0.37
VEGFC	HIF-1	0.37
GADD45B	MAPK	0.36
LEPR	HIF-1	0.36
SIPA1	Rap1	0.36
ADIPOQ	AMPK	0.35
PCDHB12	WnT	0.35
PCDHB15	WnT	0.35
PIP5K1C	Phosphatidylinositol	0.35
SNCA	AMPK	0.35
TGFB3	TGF-beta, HIF-1	0.35
TMEM173	AMPK	0.35

Table 4.1 (continued)

NOTCH4

Negatively Correlated Genes	Pathway	Pearson CC
RAN	FoxO	-0.32
CCNB1	FoxO	-0.34

L5, as tabulated in Table 4.1, indicated breast cancer genes associated with NOTCH pathway that were correlated with other signaling pathways and thus could influence the end result of those pathways. Only the genes within the Pearson Correlation Coefficient threshold of $-0.3 \geq x \geq 0.3$ were recorded in this table. The value of Pearson CC is used to determine the strength and direction of the linear statistical relation between two pathways (variables). Hence, if expression of the gene in one pathway increases, its expression in another pathway will increase and vice versa. The Pearson correlation coefficient value ranges from -1 to 1, whereas a negative value indicates negative correlation and positive value indicates positive correlation. The closer the value is to 1, the stronger the positive correlation between the two pathways (Sereno, 2021).

Chapter 5

Result

5.1 Co-expressed Gene List [L3]

The NOTCH1, NOTCH2, NOTCH3, and NOTCH4 genes co-expressed in breast cancer (derived from samples of Breast Invasive Carcinoma, TCGA, PanCancer Atlas) were derived from cBioPortal. The genes that were mentioned in both Table L5 and cBioPortal's co-expressed gene list [L3], were derived into a second list of genes L6, where the p-value and Spearman correlation coefficient for each gene was also entered. This list, L6 was filtered down to genes with p-value less than 0.05 and a Spearman CC within the threshold of $-0.4 \geq x \geq 0.4$. The filtered list of genes was labeled as L6.

The value of Spearman CC is used to determine the strength and direction of the monotonic statistical relation between two pathways (variables). Hence, if expression of the gene in one pathway increases, its expression in another pathway may increase or decrease. The Spearman correlation coefficient value ranges from -1 to 1, where a negative value indicates negative correlation and positive value indicates positive correlation. The closer the value is to -1, the stronger the negative correlation between the variables (Sereno, 2021).

5.2 Differential Expression Analysis of the Breast Cancer Dataset [L4]

This analytical procedure is carried out to compare two experimental groups of one gene expression data set. Differential expression analysis of a gene set is typically carried out to locate genes that exhibit variable expression levels after the two experimental sets are treated with different conditions. This was used to produce the list of differentially expressed genes in breast cancer. Using data from GEPIA 2, the respective Log2FC and adjusted p-value was assigned to L7, containing genes common between L6 and L4.

Log2FC measures the extent to which the expression of a gene varies across two environments, in this case, pathways.

The p-value is a statistical indication of the observed results. The lower the p-value, the greater the significance of the results. Usually, a p-value greater than 0.05 indicates that any difference in gene expression between two experimental groups is not statistically significant and has occurred only due to chance. The p-value gives the false positive rate, hence the smaller the p-value, the smaller the chances of obtaining a false positive result.

Table 5.1: NOTCH Correlated genes that are Differentially expressed in Breast Cancer (L7)

NOTCH1

Positively Correlated Genes	**Log2FC [Gepia2]**	**Spearman's Correlation (cbioportal)**	**p-value (cbioportal)**	**p-adj [Gepia2]**	**Pathway**	**Pearson CC (UALCAN)**
PLCG2	-1.2	0.58	1.65E-90	1.81E-56	VEGF, Erbb, Phosphatidylinositol	0.39
SFRP1	-3.642	0.561	1.35E-83	1.90E-85	WnT	0.35
TLE4	-1.359	0.55	9.00E-80	4.16E-81	WnT	0.4
AMOTL1	-1.155	0.542	6.38E-77	1.10E-67	Hippo	0.38
PLD1	-1.202	0.538	1.24E-75	4.75E-61	mTOR	0.41
TCF7L1	-1.58	0.515	1.56E-68	1.38E-65	WnT	0.43
CFLAR	-1.317	0.47	9.67E-56	9.91E-73	AMPK	0.36
BMP6	-2.097	0.453	1.69E-51	5.17E-145	TGF-beta	0.31
LATS2	-1.013	0.437	1.67E-47	8.96E-59	Hippo	0.32
TCF7L2	-1.302	0.436	2.85E-47	3.64E-84	WnT	0.32
AKT3	-1.357	0.428	1.20E-45	2.70E-46	MAPK, Erbb	0.35
EDN1	-1.762	0.426	4.47E-45	3.30E-71	HIF-1	0.33
LPIN1	-1.568	0.404	3.13E-40	4.02E-90	HIF-1	0.36

Table 5.1 (continued)

NOTCH1

Negatively Correlated Genes	Log2FC [Gepia2]	Spearman's Correlation (cbioportal)	p-value (cbioportal)	p-adj [Gepia2]	Pathway	Pearson CC (UALCAN)
GATA3	2.399	-0.516	1.12E-68	3.75E-31	HIF-1	-0.32

NOTCH2

Positively Correlated Genes	Log2FC	Spearman's Correlation	p-value	p-adj [Gepia2]	Pathway	Pearson CC
AMOTL1	-1.155	0.46	3.22E-53	1.10E-67	Hippo	0.35
LATS2	-1.013	0.436	2.61E-47	8.96E-59	Hippo	0.37
Negatively Correlated Genes	**Log2FC**	**Spearman's Correlation**	**p-value**	**p-adj [Gepia2]**	**Pathway**	**Pearson CC**
MIF	1.204	-0.497	5.39E-63	2.05E-72	AMPK	-0.31

NOTCH4

Positively Correlated Genes	Log2FC	Spearman's Correlation	p-value	p-adj [Gepia2]	Pathway	Pearson CC
FLT4	-1.878	0.848	3.44E-275	1.17E-137	HIF-1	0.78
PDE2A	-3.93	0.742	3.15E-174	2.96E-292	cAMP	0.67
AQP1	-2.365	0.712	1.43E-154	1.38E-116	HIF-1	0.57
ACVRL1	-1.758	0.711	4.89E-154	1.21E-124	TGF-beta	0.67
TEK	-1.566	0.706	5.93E-151	5.17E-92	HIF-1	0.58
KDR	-1.196	0.667	5.23E-129	4.69E-56	VEGF, HIF-1	0.49
CAV1	-3.01	0.653	5.66E-122	1.94E-196	TNF	0.57
NOS3	-1.24	0.644	1.61E-117	5.76E-39	VEGF, AMPK	0.58
ENG	-1.013	0.642	1.61E-116	4.33E-50	HIF-1	0.64
EDNRB	-2.707	0.629	1.82E-110	6.80E-190	HIF-1	0.49
FHL1	-4.442	0.574	5.14E-88	1.20E-212	HIF-1	0.41
EPAS1	-2.281	0.572	1.36E-87	7.54E-151	HIF-1	0.5

Table 5.1 (continued)

NOTCH4

Positively Correlated Genes	Log2FC	Spearman's Correlation	p-value	p-adj [Gepia2]	Pathway	Pearson CC
TGFBR2	-2.323	0.571	6.18E-87	7.79E-144	TGF-beta	0.37
IGF1	-2.741	0.536	3.75E-75	5.79E-118	AMPK, HIF-1	0.45
RAPGEF3	-3.41	0.517	6.85E-69	3.08E-175	Rap1, cAMP	0.58
CXCL12	-2.138	0.494	3.78E-62	7.23E-100	HIF-1	0.4
PPARG	-3.343	0.482	7.13E-59	2.15E-179	AMPK	0.38
SNCA	-1.273	0.459	5.27E-53	3.78E-92	AMPK	0.35
DUSP1	-2.629	0.457	2.49E-52	3.99E-107	MAPK	0.41
LIPE	-4.652	0.455	6.90E-52	4.69E-56	AMPK	0.47
LEPR	-2.5	0.426	5.40E-45	7.40E-170	HIF-1	0.36
ADAMTS1	-2.952	0.42	1.07E-43	1.38E-140	HIF-1	0.39
ADIPOQ	-5.615	0.42	9.31E-44	8.38E-118	AMPK	0.35
TWIST2	-3.041	0.407	5.50E-41	2.05E-167	HIF-1	0.42
Negatively Correlated Genes	**Log2FC**	**Spearman's Correlation**	**p-value**	**p-adj [Gepia2]**	**Pathway**	**Pearson CC**
RAN	1.115	-0.431	3.45E-46	1.41E-109	FoxO	-0.32
CCNB1	2.982	-0.447	5.00E-50	9.81E-209	FoxO	-0.34

The table above has been sorted based on Spearman's Correlation Coefficient (in descending order: most to least significantly correlated) and only genes within the Spearman's CC threshold of $-0.4 \geq x \geq 0.4$ have been tabulated. It also indicates if the genes are positively or negatively correlated with NOTCH pathway, On GEPIA2 database, list was obtained on the basis of Log2FC cutoff of 1 and a q-value cutoff of 0.05. The threshold for p-value was set at $p < 0.05$, so that only genes that were significantly differentially expressed between two variations can make it into the final list at Table 5.1.

5.3 List of Genes Up-regulated and Down-regulated by NOTCH Pathway

The up-regulation and down-regulation of NOTCH genes due to co-expression can be determined based on the Log2FC value for each gene. Upregulation is indicated by a positive fold change value, whereas downregulation is indicated by a negative fold change value. A specific gene, such as LPIN1 having a Log2FC value of -1.568 would imply that the expression of this gene decreases in breast cancer patients relative to healthy patients by a multiplicative factor of 2^1.568 ≈ 2.96. Similarly, CCNB1 having a Log2FC value of 2.982 would imply that the expression of this gene increases in breast cancer patients relative to healthy patients by a multiplicative factor of 2^2.982 ≈ 7.90.

Table 5.2: List of Differentially Expressed Breast Cancer Genes Up-regulated and Down-regulated by NOTCH Pathway

NOTCH1

Downregulated Gene	Log2FC	Spearman's Correlation	p-value	p-adj	Pathway	Pearson CC
SFRP1	-3.642	0.561	1.35E-83	1.90E-85	WnT	0.35
BMP6	-2.097	0.453	1.69E-51	5.17E-145	TGF-beta	0.31
EDN1	-1.762	0.426	4.47E-45	3.30E-71	HIF-1	0.33
TCF7L1	-1.58	0.515	1.56E-68	1.38E-65	WnT	0.43
LPIN1	-1.568	0.404	3.13E-40	4.02E-90	HIF-1	0.36
TLE4	-1.359	0.55	9.00E-80	4.16E-81	WnT	0.4
AKT3	-1.357	0.428	1.20E-45	2.70E-46	MAPK, Erbb	0.35
CFLAR	-1.317	0.47	9.67E-56	9.91E-73	AMPK	0.36
TCF7L2	-1.302	0.436	2.85E-47	3.64E-84	WnT	0.32
PLD1	-1.202	0.538	1.24E-75	4.75E-61	mTOR	0.41
PLCG2	-1.2	0.58	1.65E-90	1.81E-56	VEGF, Erbb, Phosphatidylinositol	0.39
AMOTL1	-1.155	0.542	6.38E-77	1.10E-67	Hippo	0.38
LATS2	-1.013	0.437	1.67E-47	8.96E-59	Hippo	0.32

Table 5.2 (continued)

NOTCH1

Upregulated Gene	Log2FC	Spearman's Correlation	p-value	p-adj	Pathway	Pearson CC
GATA3	2.399	-0.516	1.12E-68	3.75E-31	HIF-1	-0.32

NOTCH2

Downregulated Gene	Log2FC	Spearman's Correlation	p-value	p-adj	Pathway	Pearson CC
AMOTL1	-1.155	0.46	3.22E-53	1.10E-67	Hippo	0.35
LATS2	-1.013	0.436	2.61E-47	8.96E-59	Hippo	0.37
Upregulated Gene	**Log2FC**	**Spearman's Correlation**	**p-value**	**p-adj**	**Pathway**	**Pearson CC**
MIF	1.204	-0.497	5.39E-63	2.05E-72	AMPK	-0.31

NOTCH4

Downregulated Gene	Log2FC	Spearman's Correlation	p-value	p-adj	Pathway	Pearson CC
ADIPOQ	-5.615	0.42	9.31E-44	8.38E-118	AMPK	0.35
LIPE	-4.652	0.455	6.90E-52	4.69E-56	AMPK	0.47
FHL1	-4.442	0.574	5.14E-88	1.20E-212	HIF-1	0.41
PDE2A	-3.93	0.742	3.15E-174	2.96E-292	cAMP	0.67
RAPGEF3	-3.41	0.517	6.85E-69	3.08E-175	Rap1, cAMP	0.58
PPARG	-3.343	0.482	7.13E-59	2.15E-179	AMPK	0.38
TWIST2	-3.041	0.407	5.50E-41	2.05E-167	HIF-1	0.42
CAV1	-3.01	0.653	5.66E-122	1.94E-196	TNF	0.57
ADAMTS1	-2.952	0.42	1.07E-43	1.38E-140	HIF-1	0.39
IGF1	-2.741	0.536	3.75E-75	5.79E-118	AMPK, HIF-1	0.45
EDNRB	-2.707	0.629	1.82E-110	6.80E-190	HIF-1	0.49
DUSP1	-2.629	0.457	2.49E-52	3.99E-107	MAPK	0.41
LEPR	-2.5	0.426	5.40E-45	7.40E-170	HIF-1	0.36
AQP1	-2.365	0.712	1.43E-154	1.38E-116	HIF-1	0.57
TGFBR2	-2.323	0.571	6.18E-87	7.79E-144	TGF-beta	0.37
EPAS1	-2.281	0.572	1.36E-87	7.54E-151	HIF-1	0.5
CXCL12	-2.138	0.494	3.78E-62	7.23E-100	HIF-1	0.4

Table 5.2 (continued)

NOTCH4

Downregulated Gene	Log2FC	Spearman's Correlation	p-value	p-adj	Pathway	Pearson CC
FLT4	-1.878	0.848	3.44E-275	1.17E-137	HIF-1	0.78
ACVRL1	-1.758	0.711	4.89E-154	1.21E-124	TGF-beta	0.67
TEK	-1.566	0.706	5.93E-151	5.17E-92	HIF-1	0.58
SNCA	-1.273	0.459	5.27E-53	3.78E-92	AMPK	0.35
NOS3	-1.24	0.644	1.61E-117	5.76E-39	VEGF, AMPK	0.58
KDR	-1.196	0.667	5.23E-129	4.69E-56	VEGF, HIF-1	0.49
ENG	-1.013	0.642	1.61E-116	4.33E-50	HIF-1	0.64
Upregulated Gene	**Log2FC**	**Spearman's Correlation**	**p-value**	**p-adj**	**Pathway**	**Pearson CC**
CCNB1	2.982	-0.447	5.00E-50	9.81E-209	FoxO	-0.34
RAN	1.115	-0.431	3.45E-46	1.41E-109	FoxO	-0.32

This table was arranged according to the Log2FC value for each gene, where the positive values identify upregulated genes of breast cancer and negative values identify the downreguulated genes of breast cancer. For most genes it was observed that the Log2FC value was negative while the correlation coefficient was positive. In case of SFRP1, which had a Spearman's and Pearson Correlation coefficient of 0.561 and 0.35 respectively but a Log2FC value of –3.642, it indicates that while the gene was significantly positively correlated with the expression of NOTCH1 gene, its signaling was downregulated in breast cancer.

5.4 Classification of Differentially Expressed Genes in Breast Cancer under Cell-Signaling Pathways

Table 5.3: Classification of Differentially Expressed Genes in Breast Cancer under Cell-Signaling Pathways

Pathway	Gene	NOTCH Gene	Log2FC [Gepia2]	Spearman's Correlation (cbioportal)	p-value (cbioportal)	p-adj [Gepia2]	Pearson CC (UALCAN)
WnT	*TCF7L2*	NOTCH1	-1.302	0.436	2.85E-47	3.64E-84	0.32
	TCF7L1	NOTCH1	-1.58	0.515	1.56E-68	1.38E-65	0.43
	TLE4	NOTCH1	-1.359	0.55	9.00E-80	4.16E-81	0.4
	SFRP1	NOTCH1	-3.642	0.561	1.35E-83	1.90E-85	0.35
MAPK	*AKT3*	NOTCH1	-1.357	0.428	1.20E-45	2.70E-46	0.35
	DUSP1	NOTCH4	-2.629	0.457	2.49E-52	3.99E-107	0.41
Erbb	*AKT3*	NOTCH1	-1.357	0.428	1.20E-45	2.70E-46	0.35
	PLCG2	NOTCH1	-1.2	0.58	1.65E-90	1.81E-56	0.39
TGF-β	*BMP6*	NOTCH1	-2.097	0.453	1.69E-51	5.17E-145	0.31
	TGFBR2	NOTCH4	-2.323	0.571	6.18E-87	7.79E-144	0.37
	ACVRL1	NOTCH4	-1.758	0.711	4.89E-154	1.21E-124	0.67
Hippo	*LATS2*	NOTCH1	-1.013	0.437	1.67E-47	8.96E-59	0.32
	AMOTL1	NOTCH1	-1.155	0.542	6.38E-77	1.10E-67	0.38
	LATS2	NOTCH2	-1.013	0.436	2.61E-47	8.96E-59	0.37
	AMOTL1	NOTCH2	-1.155	0.46	3.22E-53	1.10E-67	0.35
VEGF	*PLCG2*	NOTCH1	-1.2	0.58	1.65E-90	1.81E-56	0.39
	NOS3	NOTCH4	-1.24	0.644	1.61E-117	5.76E-39	0.58
	KDR	NOTCH4	-1.196	0.667	5.23E-129	4.69E-56	0.49
TNF	*CAV1*	NOTCH4	-3.01	0.653	5.66E-122	1.94E-196	0.57
HIF-1	*LPIN1*	NOTCH1	-1.568	0.404	3.13E-40	4.02E-90	0.36
	EDN1	NOTCH1	-1.762	0.426	4.47E-45	3.30E-71	0.33
	GATA3	NOTCH1	2.399	-0.516	1.12E-68	3.75E-31	-0.32
	TWIST2	NOTCH4	-3.041	0.407	5.50E-41	2.05E-167	0.42
	ADAMTS1	NOTCH4	-2.952	0.42	1.07E-43	1.38E-140	0.39

Table 5.3 (continued)

Pathway	Gene	NOTCH Gene	Log2FC [Gepia2]	Spearman's Correlation (cbioportal)	p-value (cbioportal)	p-adj [Gepia2]	Pearson CC (UALCAN)
HIF-1	*CXCL12*	NOTCH4	-2.138	0.494	3.78E-62	7.23E-100	0.4
	IGF1	NOTCH4	-2.741	0.536	3.75E-75	5.79E-118	0.45
	EPAS1	NOTCH4	-2.281	0.572	1.36E-87	7.54E-151	0.5
	FHL1	NOTCH4	-4.442	0.574	5.14E-88	1.20E-212	0.41
	EDNRB	NOTCH4	-2.707	0.629	1.82E-110	6.80E-190	0.49
	ENG	NOTCH4	-1.013	0.642	1.61E-116	4.33E-50	0.64
	KDR	NOTCH4	-1.196	0.667	5.23E-129	4.69E-56	0.49
	TEK	NOTCH4	-1.566	0.706	5.93E-151	5.17E-92	0.58
	AQP1	NOTCH4	-2.365	0.712	1.43E-154	1.38E-116	0.57
	FLT4	NOTCH4	-1.878	0.848	3.44E-275	1.17E-137	0.78
FoxO	*CCNB1*	NOTCH4	2.982	-0.447	5.00E-50	9.81E-209	-0.34
	RAN	NOTCH4	1.115	-0.431	3.45E-46	1.41E-109	-0.32
Phosphatidyl -inositol	*PLCG2*	NOTCH1	-1.2	0.58	1.65E-90	1.81E-56	0.39
cAMP	*RAPGEF3*	NOTCH4	-3.41	0.517	6.85E-69	3.08E-175	0.58
AMPK	*CFLAR*	NOTCH1	-1.317	0.47	9.67E-56	9.91E-73	0.36
	MIF	NOTCH2	1.204	-0.497	5.39E-63	2.05E-72	-0.31
	ADIPOQ	NOTCH4	-5.615	0.42	9.31E-44	8.38E-118	0.35
	LIPE	NOTCH4	-4.652	0.455	6.90E-52	4.69E-56	0.47
	SNCA	NOTCH4	-1.273	0.459	5.27E-53	3.78E-92	0.35
	PPARG	NOTCH4	-3.343	0.482	7.13E-59	2.15E-179	0.38
	IGF1	NOTCH4	-2.741	0.536	3.75E-75	5.79E-118	0.45
	NOS3	NOTCH4	-1.24	0.644	1.61E-117	5.76E-39	0.58
mTOR	*PLD1*	NOTCH1	-1.202	0.538	1.24E-75	4.75E-61	0.41
Rap1	*RAPGEF3*	NOTCH4	-3.41	0.517	6.85E-69	3.08E-175	0.58

The table above classified genes on the basis of the signal transduction processes they are associated with (except for NOTCH). It also mentions the NOTCH gene they are usually correlated with in breast cancer.

Table 5.4 below shows pathways containing genes that are both highly significantly correlated with NOTCH signaling (those with high Spearman correlation coefficient values) and substantially regulated in breast cancer (significantly high or low Log2FC value):

Table 5.4: Genes that are highly significantly correlated with NOTCH signaling

Pathway	Pathway Function in Cancer	Genes	Log2FC	Spearman's CC
WnT (Wingless-related integration site) Pathway	Cell renewal, cell migration, cell proliferation and cell differentiation.	*SFRP1*	-3.642	0.561
Erbb (Erythroblastic leukemia viral oncogene homologue) Pathway	Cell proliferation, cell survival, cell migration, cell growth, cell differentiation, apoptosis, and cell motility.	*PLCG2*	-1.2	0.58
TGF-β (Transforming growth factor-beta) Pathway	Anti-proliferative effect, cell growth, cell differentiation, cell migration, apoptosis.	*TGFBR2*	-2.323	0.571
		ACVRL1	-1.758	0.711
VEGF (Vascular endothelial growth factor) Pathway	Endothelial cell growth, angiogenesis, cell migration, and cell survival.	*PLCG2*	-1.2	0.58
		NOS3	-1.24	0.644
		KDR	-1.196	0.667
TNF (Tumor necrosis factor) Pathway	Tumor promoter, cell proliferation, cell invasion, metastasis, angiogenesis, cell survival and apoptosis.	*CAV1*	-3.01	0.653
HIF-1 (Hypoxia-inducible factor-1) Pathway	Tumor angiogenesis	*CXCL12*	-2.138	0.494
		IGF1	-2.741	0.536
		EPAS1	-2.281	0.572
		FHL1	-4.442	0.574
		EDNRB	-2.707	0.629
		ENG	-1.013	0.642
		KDR	-1.196	0.667
		TEK	-1.566	0.706
		AQP1	-2.365	0.712
		FLT4	-1.878	0.848
Phosphatidylinositol Pathway	Cell proliferation, cell growth, apoptosis, cell invasion, cell survival.	*PLCG2*	-1.2	0.58

Table 5.4 (continued)

Pathway	Pathway Function in Cancer	Genes	Log2FC	Spearman's CC
AMPK (AMP-activated protein kinase) Pathway	Tumor suppressor, cell-cycle arrest	*NOS3*	-1.24	0.644
mTOR (mammalian target of rapamycin) Pathway	Cell survival, cell growth and protein synthesis	*PLD1*	-1.202	0.538

The table above narrowed down the contents of Table 5.3 and only comprises of the gene(s) from each pathway that have both high Spearman CC value (preferably close to the threshold $-0.5 \geq x \geq 0.5$) and significant Log2FC value (highly negative or highly positive)

5.5 Gene Enrichment Analysis of Differentially Expressed Genes in Breast Cancer

Gene Set Enrichment analysis for a gene set (also known as over-representation analysis) is an essential bioinformatics tool that can help recognize the biological pathways that are common or over-represented within the gene list more than what would be expected by chance (Chicco & Agapito, 2022). Hence, one can determine the specific genes from the gene set that are commonly involved in different biological pathways.

The results indicate the biological pathways in which the genes from the set are involved, the total number of genes usually involved in the pathway, the total number of genes from the input set that are involved in the pathway (nGenes), the Fold Enrichment for the pathway, the FDR (False Discovery Rate) of the Fold Enrichment result and lastly, the symbols of the involved genes from the input gene set.

The fold enrichment value merely refers to the number of times more something occurred than what was predicted by chance; for instance, a fold enrichment of 3 implies it happened thrice as frequently as the random expectation.

The FDR value indicates the significance of the enrichment results. An FDR < 0.05 would imply that the fold enrichment outcome is statistically significant and not due to chance.

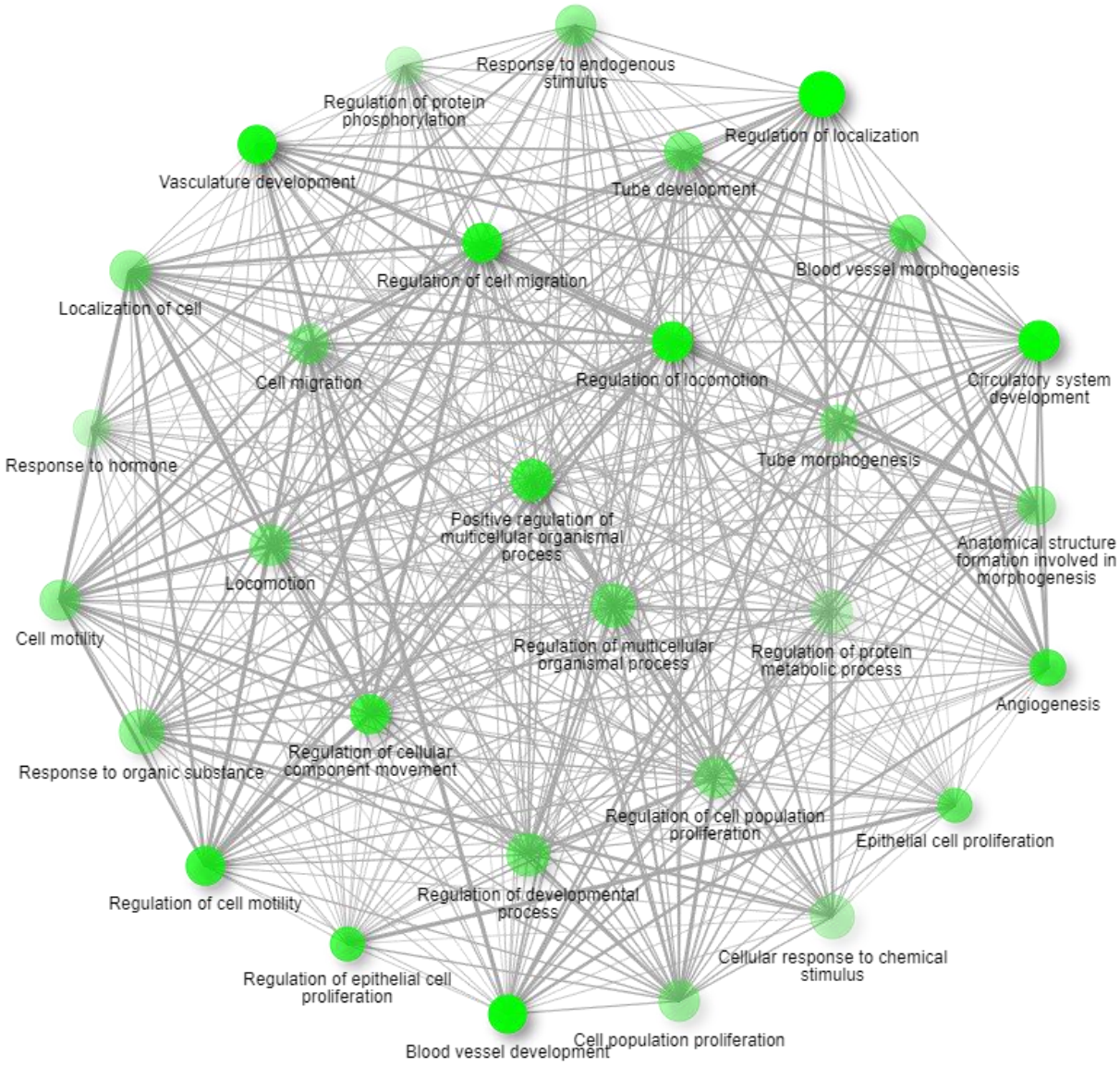

Figure 5.1: Network between enriched pathways (*ShinyGO V0.741: Gene Ontology Enrichment Analysis + More*, 2022)

The list of differentially expressed genes in *Table 5.1* were subjected to enrichment analysis in ShinyGO v0.741: Gene Ontology Enrichment Analysis + more database. Initially the Ensembl ID (e.g. *PLCG2*: ENSG00000197943) for each gene was derived from Ensembl Database and the gene set enrichment result was as follows:

Table 5.5: Gene Enrichment Analysis of Differentially Expressed Genes in Breast Cancer

Pathway	Pathway Genes	nGenes	Fold Enrichment	Enrichment FDR	Genes
Epithelial cell proliferation	409	16	21.75061125	1.72E-15	*CFLAR, IGF1, FLT4, SFRP1, CAV1, ENG, GATA3, CXCL12, TEK, KDR, PPARG, EDNRB, ACVRL1, TCF7L2, BMP6, AKT3*
Angiogenesis	544	18	18.39705882	2.03E-16	*FLT4, EDN, RAPGEF3, SFRP, CAV, ENG, EPAS1, LEPR, TEK, KDR, PPARG, ACVRL1, ADAMTS1, TGFBR2, NOS3, AMOTL1, AQP1, AKT3*
Regulation of cell motility and cell migration	1010	22	12.11089109	7.32E-17	*IGF1, FLT4, EDN1, SFRP1, CAV1, ENG, GATA3, CXCL12, DUSP1, TEK, KDR, PPARG, ACVRL1, ADAMTS1, TGFBR2, NOS3, AMOTL1, ADIPOQ, PLCG2, AKT3, MIF, TWIST2*
Regulation of locomotion	1056	23	12.10984848	1.04E-17	*IGF1, FLT4, EDN1, SFRP1, CAV1, ENG, GATA3, CXCL12, DUSP1, TEK, KDR, PPARG, ACVRL1, SNCA, ADAMTS1, TGFBR2, NOS3, AMOTL1 ADIPOQ, PLCG2, AKT3, MIF, TWIST2*
Vasculature and Circulatory system development	1127	24	11.84028394	3.68E-18	*CFLAR, IGF1, FLT4, EDN1, RAPGEF3, SFRP1, CAV1, ENG, GATA3, EPAS1, LEPR, TEK, KDR, PPARG, CCNB1, ACVRL1, TCF7L2, ADAMTS1, TGFBR2, NOS3, AMOTL1, PDE2A, AQP1, AKT3*
Response to hormone	998	19	10.58517034	1.52E-13	*CFLAR, EDN1, SFRP1, CAV1, ENG, GATA3, CXCL12, LEPR, DUSP1, TEK, PPARG, LPIN1, EDNRB, LATS2, BMP6, TGFBR2, NOS3, ADIPOQ, AQP1*

Table 5.5 (continued)

Pathway	Pathway Genes	nGenes	Fold Enrichment	Enrichment FDR	Genes
Regulation of protein phosphorylation	1167	20	9.528706084	1.52E-13	*IGF1, FLT4, EDN1, RAPGEF3, SFRP1, CAV1, ENG, LEPR, DUSP1, TEK, KDR, CCNB1, EDNRB, ACVRL1, SNCA, LATS2, BMP6, TGFBR2, ADIPOQ, MIF*
Regulation of cell population proliferation	1817	24	7.343973583	2.10E-14	*CFLAR, IGF1 ,FLT4, EDN1, SFRP1, CAV1, ENG, GATA3, CXCL12, TEK, KDR, PPARG, CCNB1, EDNRB , ACVRL1, TCF7L2 , BMP6, ADAMTS1, TGFBR2, NOS3, ADIPOQ, AQP1, AKT3, MIF*
Cellular response to chemical and organic stimulus	3680	30	4.532608696	1.01E-13	*CFLAR, IGF1, FLT4, EDN1, RAPGEF3, SFRP1, CAV1, TLE4, ENG ,GATA3, CXCL12, EPAS1, LEPR, DUSP1, TEK, KDR, PPARG, CCNB1, LPIN1, EDNRB, ACVRL1, SNCA, TCF7L2, LATS2, BMP6, TGFBR2, NOS3, ADIPOQ, PDE2A, PLCG2, AQP1, MIF*

The table indicates the different biological processes that can be affected by the gene list that were previously determined to be differentially expressed in breast cancer (Table 5.1). This can be used to further analyze which of these processes are related to breast cancer tumor formation, growth, proliferation and survival. Figure 5.2 below schematically summarizes the result.

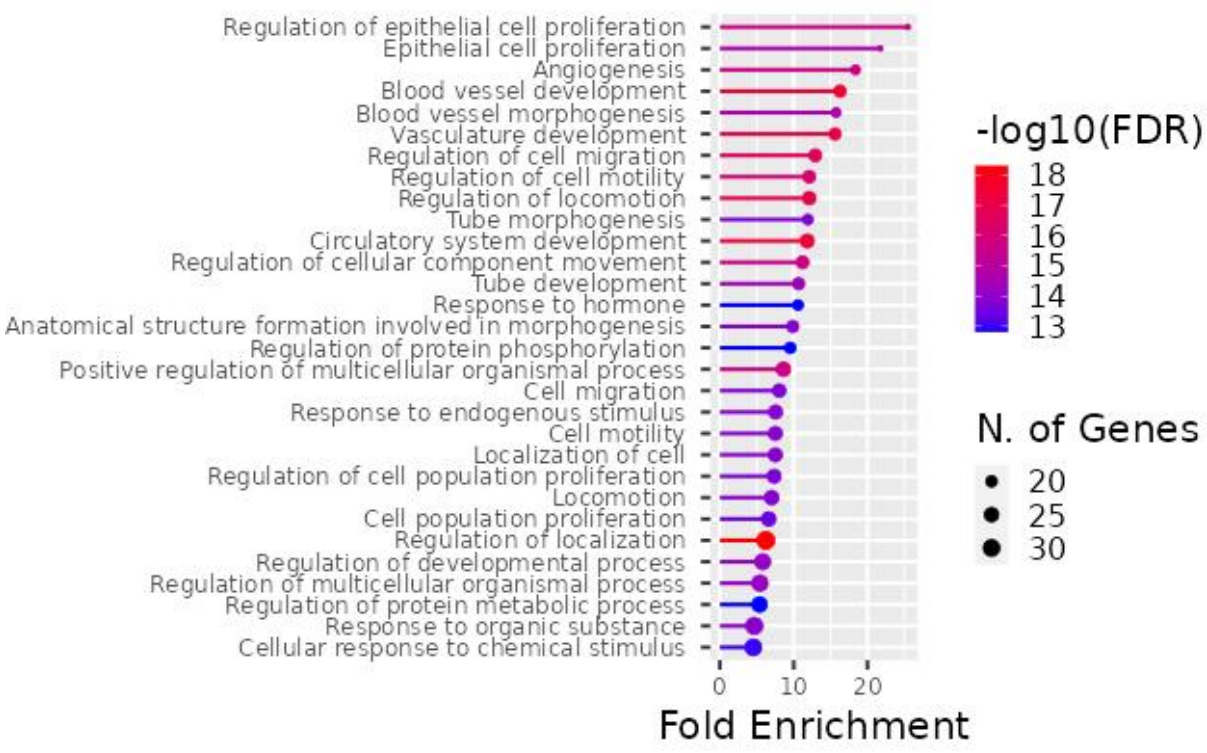

Figure 5.2: Gene Enrichment Analysis of Differentially Expressed Genes in Breast Cancer (*ShinyGO V0.741: Gene Ontology Enrichment Analysis + More*, 2022)

As established before, cancer progression is primarily linked to the 10 Hallmarks of Cancer:

1. Sustained proliferative signaling
2. Evading growth suppressors
3. Tissue invasion and metastasis
4. Triggering angiogenesis
5. Facilitating replicative immortality
6. Evading apoptosis
7. Avoiding immune destruction
8. Reforming cellular energy and metabolic processes
9. Genome Instability and Mutation
10. Tumor-promoting inflammation

Other than these reasons, some other factors facilitating the growth and invasion of malignant tumor cells are: cell motility, cell migration, locomotion, call localization, ability to respond to external stimuli etc. Based on this, the rest of the paper will focus only on the following pathways and their related genes:

a) Regulation of epithelial cell proliferation
b) Regulation of localization
c) Angiogenesis
d) Blood vessel development
e) Regulation of cell migration
f) Regulation of cell motility
g) Regulation of locomotion
h) Regulation of cellular component movement
i) Response to hormone
j) Response to endogenous stimulus

Table 5.6 has been constructed to display the presence-absence of each differentially expressed gene in the enriched pathways. This information was additionally tabulated into *Figure 5.2,* where only pathways with a fold enrichment value greater than 6 were considered.

Table 5.6: Genes involved Pathways and processes related to Breast Cancer

Gene	**Correlated NOTCH Gene**	Epithelial cell proliferation	Angiogenesis	Regulation of cell motility and migration	Regulation of locomotion	Vasculature and Circulatory system development	Response to hormone	Regulation of protein phosphorylation	Regulation of cell population proliferation	Cellular response to chemical and organic stimulus
PLCG2	NOTCH1			✓	✓					✓
SFRP1	NOTCH1	✓	✓	✓	✓	✓	✓	✓	✓	✓
TLE4	NOTCH1									✓
AMOTL1	NOTCH1, NOTCH2		✓	✓	✓	✓				
PLD1	NOTCH3									
TCF7L1	NOTCH1									
CFLAR	NOTCH1	✓				✓	✓		✓	✓
BMP6	NOTCH3	✓					✓	✓	✓	✓
LATS2	NOTCH1, NOTCH2						✓	✓		✓
TCF7L2	NOTCH1	✓				✓			✓	✓
AKT3	NOTCH1	✓	✓	✓	✓	✓			✓	
EDN1	NOTCH1		✓	✓	✓	✓	✓	✓	✓	✓
LPIN1	NOTCH1						✓			✓
GATA3	NOTCH1	✓		✓	✓	✓	✓		✓	✓
MIF	NOTCH2			✓	✓			✓	✓	✓
FLT4	NOTCH4	✓	✓	✓	✓	✓		✓	✓	✓
PDE2A	NOTCH4					✓				✓
AQP1	NOTCH4		✓			✓	✓		✓	✓
ACVRL1	NOTCH4	✓	✓	✓	✓	✓		✓	✓	✓
TEK	NOTCH4	✓	✓	✓	✓	✓	✓	✓	✓	✓
KDR	NOTCH4	✓	✓	✓	✓	✓		✓	✓	✓
CAV1	NOTCH4	✓	✓	✓	✓	✓	✓	✓	✓	✓
NOS3	NOTCH4		✓	✓	✓	✓	✓		✓	✓
ENG	NOTCH4	✓	✓	✓	✓	✓	✓	✓	✓	✓
EDNRB	NOTCH4	✓					✓	✓	✓	✓
FHL1	NOTCH4									

Table 5.6 (continued)

Gene	**Correlated NOTCH Gene**	Epithelial cell proliferation	Angiogenesis	Regulation of cell motility and migration	Regulation of locomotion	Vasculature and Circulatory system development	Response to hormone	Regulation of protein phosphorylation	Regulation of cell population proliferation	Cellular response to chemical and organic stimulus
EPAS1	NOTCH4		✓			✓				✓
TGFBR2	NOTCH4		✓	✓	✓	✓	✓	✓	✓	✓
IGF1	NOTCH4	✓	✓	✓	✓	✓		✓	✓	✓
RAPGEF3	NOTCH4		✓			✓		✓		✓
CXCL12	NOTCH4	✓		✓	✓		✓		✓	✓
PPARG	NOTCH4	✓	✓	✓	✓	✓	✓		✓	✓
SNCA	NOTCH4				✓			✓		✓
DUSP1	NOTCH4			✓	✓		✓	✓		✓
LIPE	NOTCH4									
LEPR	NOTCH4		✓			✓	✓	✓		✓
ADAMTS1	NOTCH4		✓	✓	✓	✓			✓	
ADIPOQ	NOTCH4			✓	✓		✓	✓	✓	✓
TWIST2	NOTCH4			✓	✓					
RAN	NOTCH4									
CCNB1	NOTCH4					✓		✓	✓	✓

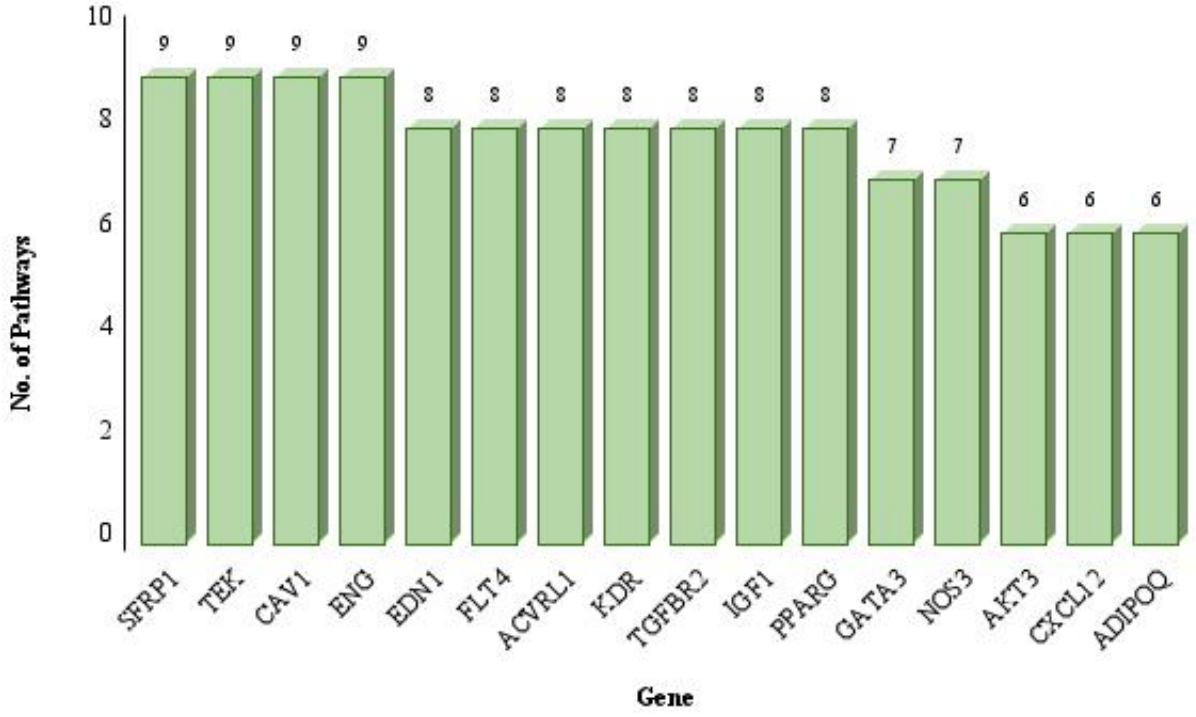

Figure 5.3: Schematic representation of the number of pathways associated with each differentially expressed gene

For each gene, the number of pathways it is associated with, is determined and tabulated in *Figure 5.2.* From this table, it can be derived that the genes involved in 6 or more pathways have significant potential to affect the biological processes of the body if they are considered as druggable targets in breast cancer drug development.

According to the analysis so far, the potential genes are: *SFRP1, TEK, CAV1, ENG, EDN1, FLT4, ACVRL1, KDR, TGFBR2, IGF1, PPARG, GATA3, NOS3, AKT3, CXCL12, ADIPOQ*

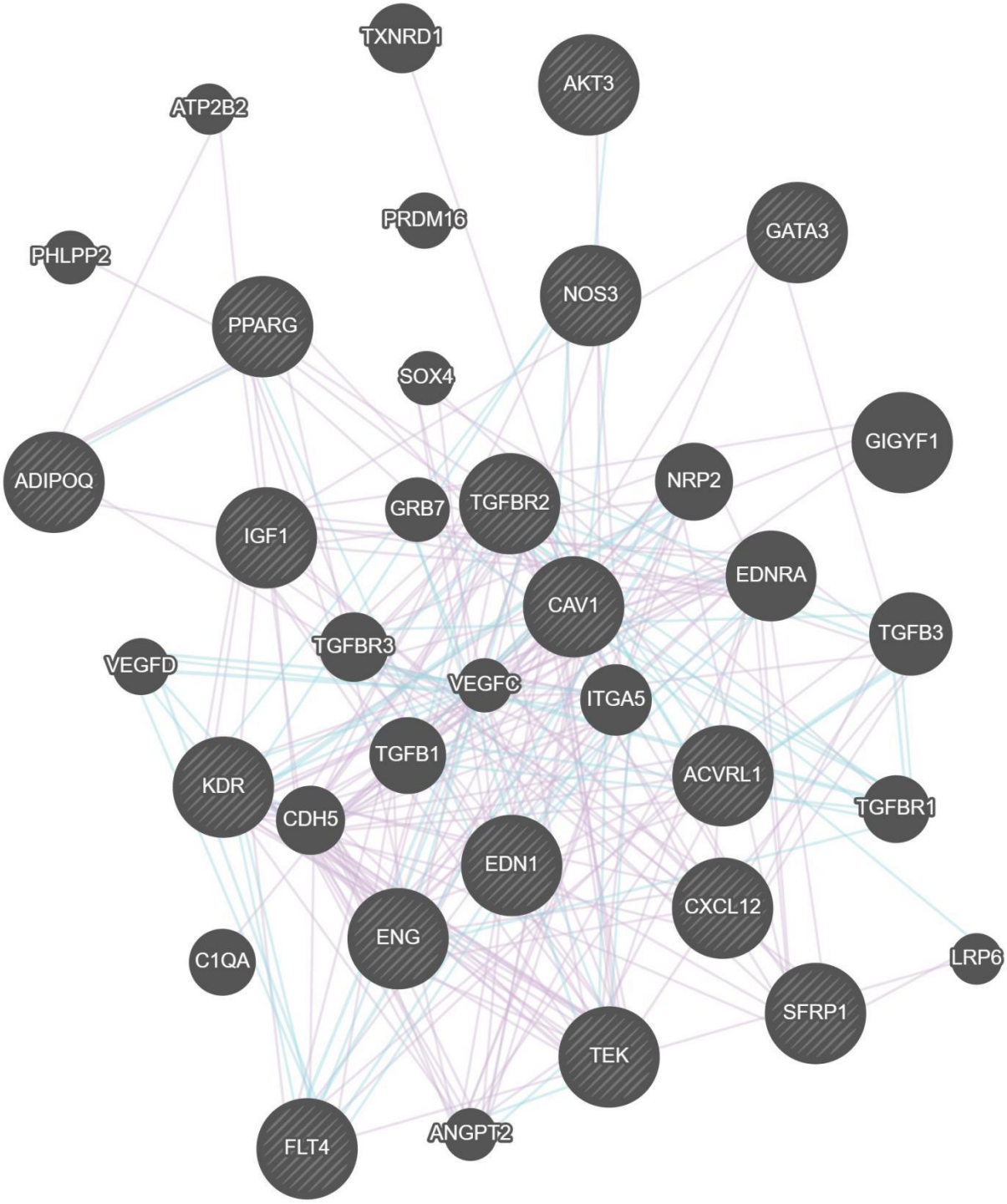

Figure 5.4: Co-expression (pink) and Pathway interaction (blue) between potential genes (*GeneMANIA*, 2024)

5.6 Functional SNPs of Potential Genes

The SVM score indicates the type of variant of the given gene that was formed due to single nucleotide polymorphism. A positive SVM value indicates formation of a non-deleterious variant that alters protein function, whereas a negative SVM value indicates the formation of a deleterious variant which is more likely to affect protein stability. A deleterious variant is formed by deleterious mutation, which makes the cells or organism more prone to genetic disorders whereas a non-deleterious variant is usually not a causative factor behind genetic diseases. The greater the magnitude of the SVM score, the greater the accuracy of the result (variant type). In most cases SVM scores in the range –0. 5 ≥ x ≥ 0.5 signifies an accurate result.

Table 5.7*:* SNP Variants of Potential Genes

Gene	snp	snp_id	svm profile	Gene	snp	snp_id	svm profile
TEK	I148T	35969327	2.36	*NOS3*	Q112R	3918166	2.77
	V600L	35030851	2.08		R885M	3918201	2.6
	A226V	35814893	1.61		E298D	1799983	2.29
	A724T	4631561	1.09		R665H	7792133	1.52
	L752F	45510599	1.05		Q982L	3918234	0.73
	N452D	45607336	0.88		V827M	3918232	0.71
	V486I	1334811	0.78		R255Q	41377046	0.54
	T391I	**34032300**	**-1.54**		**C991S**	**6947833**	**-0.72**
PPARG	P113Q	1800571	1.89		**R372C**	**41508746**	**-3.57**
	R316H	**28936407**	**-0.54**	*TGFBR2*	T340M	34833812	0.99
ENG	H366D	1800956	2.18		T315M	34833812	0.94
	T5M	35400405	1.6		V439A	1050833	0.69
	G191D	**41322046**	**-0.87**		V464A	1050833	0.67
	P469S	**11545662**	**-0.91**		**L308P**	**28934568**	**-2.79**
	T295R	12042	-1.89		**L333P**	**28934568**	**-2.84**

Table 5.7 (continued)

Gene	snp	snp_id	svm profile	Gene	snp	snp_id	svm profile
	H868Y	35171798	1.57		V417L	35011589	0.68
	P954S	34255532	1.46	*GATA3*	V416L	35011589	0.65
	N527S	35874891	1.42		G242S	11567901	0.56
	T494A	307826	1.34	*CXCL12*	**W78R**	**11541200**	**-1.54**
	P954S	34255532	1.19		**W78R**	**11541200**	**-1.45**
FLT4	N527S	35874891	1.15		T689M	34038364	2.14
	T494A	307826	1.04		H472Q	1870377	1.87
	N149D	34221241	0.7		V952I	13129474	1.42
	H890Q	448012	0.59		R349K	1824302	1.15
	H868Y	35171798	0.59		C482R	34231037	0.74
	R1189C	744282	0.58		A772T	1062832	0.71
	C344Y	**28936688**	**-2.56**	*KDR*	S1347T	1139777	0.5
	M376R	**28936399**	**-2.62**		**D392N**	**2034964**	**-0.61**
ACVRL1	**I245N**	**1804508**	**-2.85**		**K835N**	**1139775**	**-0.8**
	G211D	**28936687**	**-3.07**		**V297I**	**2305948**	**-0.82**
	R374W	**28936401**	**-3.76**		**R842C**	**41469552**	**-1.32**
IGF1	A115T	17884626	1.91		**R787G**	**1139774**	**-1.37**
ADIPOQ	**G54V**	**13061862**	**-3.74**		**V848E**	**1139776**	**-3.37**
	Y111H	17366743	1.61				

It is important to understand that snp variants with both highly positive and highly negative svm score are important elemnents in this study. The non-delereious variants with positive score should be seen as potential and desired replacements for parent genes that generally promote tumor progression and migration. On the other hand, the deleterious variants with a negative score should be seen as targets that must be inhibied to stop cancer development.

The *Table 5.7* demonstrates the possible gene variations of the potential genes. The significance of each SNP variation on biological pathways can be understood by observing their svm score, based on which it was concluded that the deleterious variants (highly negative svm score) of the genes *TEK, ENG, ACVRL1, KDR, CXCL12, NOS3, TGFBR2, and*

ADIPOQ are most likely to cause mutations that may eventually lead to genetic disorders. On the other hand, the non-deleterious variants (highly positive svm score) of TEK, ENG, KDR, NOS3, IGF1, and PPARG have a significant effect on the function or stability of the protein generally coded by the parent gene but are less likely to induce genetic diseases. Accordingly, non-deleterious variants of *TEK, ENG, KDR, NOS3, IGF1, and PPARG* are our genes of choice as their expression can alter their general activity of the parent gene without exerting any mutagenic effect on body. On the contrary, some variants of *TEK, ENG, KDR, NOS3, IGF1, PPARG, ACVRL1, ADIPOQ, TGFBR2* make the cells more prone to mutation when their respective proteins are expressed and thus, these polymorphisms need to be inhibited. This was determined by examining the svm score of the variants of each gene. Although any variant in accordance with the specified threshold $-0.5 \geq x \geq 0.5$ is a promising one, for this study, the variants with highest positive svm score and lowest negative svm score were selected as best possible targets.

TEK: It is associated with the synthesis of TEK receptor tyrosine kinase protein which is an endothelial cell specific receptor for the glycoproteins: ANGPT1, ANGPT2 and ANGPT4. Amongst biological pathways, *TEK* gene is often linked to the regulation of angiogenesis, vasculogenesis, cell migration, cell adhesion, cell proliferation, cell survival, cell invasion and vascular inactivation. TEK exerts inhibitory action on angiogenesis, while also enhancing adhesion between migratory cells (Bateman et al., 2022). This gene is commonly overexpressed in breast cancer to facilitate the proliferation, growth and invasion of carcinogenic cells to distant regions of the body (Lobato, 2024). Hence, inhibition of the *TEK* gene can largely contribute to the control of tumor growth in breast cancer patients. Therefore, it can be regarded as a promising target for breast cancer drug development.

According to *Table 5.7,* the two most promising non-deleterious variants of *TEK* gene, I148T and V600L have the svm scores 2.36 and 2.08 respectively, which indicates that these variants will code for a protein that has altered function as compared to the protein coded by original *TEK* gene. From this information, it can be derived that external chemical stimulus such as anti-cancer drugs that can induce formation of I148T and V600L variants of *TEK* gene, will successfully prevent TEK receptor tyrosine kinase protein from functioning efficiently and subsequently prevent it from facilitating angiogenesis, vasculogenesis, migration, adhesion, survival and invasion of tumor cells, thereby limiting the advancement of breast cancer cells. On the contrary, drugs that can induce formation of the deleterious variant, T391I, which has an svm score of -1.54, will code for a protein that promotes cellular mutation and tumorigenesis. Accordingly, drugs must also possess the ability to inhibit single-nucleotide polymorphism of TEK into T391I.

ENG: This gene codes for the vascular glycoprotein known as type I membrane protein [44], which is an important regulator of vascular angiogenesis, morphogenesis, metastasis and migratory action of cells. As a result, ENG plays an integral part in maintaining the integrity of vasculature (Bateman et al., 2022). Its overexpression is generally noted in tumorous vessels and endothelial cells where it promotes angiogenesis and any error in its regulation or expression may lead to development of tumor. On the contrary, ENG expression has tumor suppressive action on tumorous cells, where it limits malignant cell proliferation, invasion and motility (Lobato, 2022). Due to its dual effect on body cells, this makes ENG a clinically difficult gene to regulate via external substances such as drugs but owing to its significant influence on both tumor cells and endothelial vascular cells, it is crucial to regard *ENG* as a potential target for anticancer drugs.

According to *Table 5.7,* the H366D variant of *ENG* gene with a svm scores 2.18 will code for a protein that has altered function or stability as compared to the protein coded by original *ENG* gene. From this information, it can be derived that external chemical stimulus such as anti-cancer drugs that can induce formation of H366D variant of *ENG* gene, will form a faulty or non-functional protein. This will stop the tumor suppressive activity of ENG on tumorous cells while also preventing ENG-induced angiogenesis in endothelial cells. On the other hand, polymorphism of *ENG* into its deleterious variant T295R, with a negative svm score of -1.89, will express a protein with possibly mutagenic properties. Therefore, drugs must focus on enhancing the expression of original *ENG* gene in tumor cells, rather than inducing variant formation via polymorphism.

KDR: Kinase insert domain receptor (*KDR*) codes for a VEGF receptor which is a type III tyrosine kinase receptor for VEGFA, VEGFC and VEGFD. This protein has a significant role in VEGF-induced cell proliferation, differentiation, migration, morphogenesis and survival. It is a notably influential gene due to its ability to affect several other cell-signaling pathways such as MAPK1/ERK2, MAPK3/ERK1, AKT1 pathway (Bateman et al., 2022). Additionally, this receptor's expression is also extensively enhanced during endothelial tumor angiogenesis which further proves its importance in tumor advancement. Studies have established that suppression of *KDR* gene with specific inhibitors can successfully reduce the development of tumor (Lobato, 2015).

From *Table 5.7,* T689M and H472Q can be determined as the non-deleterious variants of *KDR* gene with the svm scores 2.14 and 1.87 respectively. This signifies that these variants will code for a protein that has modified function as compared to the protein coded by original *KDR* gene. As KDR has a prominent role in tumor angiogenesis, migration, differentiation and survival, an alteration in its expression will actively reduce the

development and growth of malignant cells. Hence, inhibitory effect of drugs on KDR gene or drug-induced polymorphism of *KDR* into T689M and H472Q can help control cancer cell proliferation and invasion. The table also highlights the presence of a deleterious variant of *KDR*, V848E with a highly negative svm score of -3.37. This confirms that single-nucleotide polymorphism of *KDR* into V848E and the resulting expression of this gene will code for a protein with altered function and stability and also contribute to enhanced tumor proliferation. Hence, a drug targeting *KDR* gene must prevent its polymorphism into V848E.

NOS3: Nitric Oxide Synthase 3 (*NOS3*) gene is primarily involved in formation of nitric oxide in cells but lately data has proved its potential effect on cellular growth and other biological processes which makes it a viable candidate for development of anti-cancer drugs. Recent studies have linked NOS3 to enhanced development of cancerous tumors where it has been implicated in promotion of cellular proliferation, migration, invasion, angiogenesis. Moreover, this gene aids tumor cells in developing anti apoptotic properties while suppressing the immune surveillance system of the body. Further relation of *NOS3* to poor prognosis in cancer and its subsequent correlation to several other pathways makes it a gene of interest (Zou et al., 2021).

Data from *Table 5.7* identifies Q112R and R885M, with svm scores 2.77 and 2.6 respectively, as promising non-deleterious variants of NOS3 gene. Thus, single nucleotide polymorphism of *NOS3* will form these variants that will have variable protein expression as compared to the original gene. *NOS3* has been linked to cancer cells possessing anti-apoptotic properties and inhibition of this gene or alteration in its protein's structure, function or stability via polymorphism will significantly aid in regulation of the exponential progression of breast cancer cells. At the same time, it is important to note the presence of a deleterious variant of *NOS3* in *Table 5.7*, with a striking high negative svm score 0f -3.57. Expression of

the R372C variant will enhance the tumerigenic property of cancer cells and therefore prevention of this variant's formation must be taken into consideration when developing an anti-cancer drug.

IGF1: Insulin Like Growth Factor 1 (*IGF1*) codes for a protein whose activity resembles that of insulin and is generally linked with biological processes related to cellular growth, repair and development (Bateman et al., 2022). Its active role in several signaling pathways such as AKT1, PI3K-AKT/PKB, Ras-MAPK, MAPK3/ERK1 and MAPK1/ERK2 makes it a versatile gene to target for anti-cancer drug development. Failure to efficiently regulate this gene can create disruptions in the pathways it is a part of and subsequently result in anti-apoptotic or tumor growth promoting effects (Lobato, 2024). Therefore, drug development based on *IGF1* gene must focus on its controlled expression in every pathway it is associated with.

Data demonstrated in *Table 5.7* identifies A115T as a plausible non-deleterious variant of IGF1, with a svm score of 1.91. This indicates that the variant is capable of expressing a protein which is a functionally altered version of the protein originally expressed by *IGF1* gene. Subsequent expression of this altered protein will stop the processes otherwise coordinated and enhanced by the *IGF1*-expressed protein. Drugs or chemicals that can successfully inhibit *IGF1* or activate polymorphism of the gene will be able to curb the cellular growth, repair and development related function of IGF1. As a result, the progression of cancerous tumor development can be restrained.

PPARG: Peroxisome Proliferator Activated Receptor Gamma gene is primarily involved in the differentiation of adipocytes and regulatioon of homeostasis (Bateman et al., 2022). Different studies have established the significant role of this gene in breast cancer, where it is

usually downregulated. This further proves that overexpression of *PPARG* gene can create anti-tumor effect. Additionally, it plays a crucial immunological role in regulation of cell metabolism, cancer cell development, growth and progression (Li et al., 2023). The anti-tumerigenic activity of *PPARG* makes it a promising target for development of anti-cancer drugs.

Table 5.7 can be used to identify P113Q as a non-deleterious variant of *PPARG*, with a svm score of 1.89. It can be understood that the variant will code for a protein with different function and stability than the protein coded by the original gene. As the activation of this gene is important to slow down cancer progression, drug development must focus upon upregulation *PPARG* expression and inhibit its polymorphism into the non-deleterious variant, P113Q. At the same time, it is equally important to consider its deleterious variant, R316H, whose expression can give rise to further mutagenic disorders and hence, a possible anti-cancer drug must resist inducing the formation or expression of R316H.

ACVRL1: Activin A receptor like type 1 (*ACVRL1*) codes for a type I cell-surface receptor ALK1, which is a serine-threonine kinase. It codes for a protein that has prominent function in the regulation of blood vessel growth, which explains its high expression in endothelial cells (Bateman et al., 2022). Through the TGF-β signaling pathway, ACVRL1 prevents exponential growth of endothelial cells of unwanted morphology. In contrast, it has been reported to have been a contributory factor in cancer neoangiogenesis, hence facilitating the proliferation, differentiation and migration of tumor cells to distant locations of body (Lobato, 2014). All these features have made *ACVRL1* the subject of anti-cancer drug development.

Data from *Table 5.7* shows that all possible snp variants of *ACVRL1* are deleterious variants with a negative svm score. As a result, activation of protein normally coded by these variants can promote mutation, tumor cell differentiation and migration in breast cancer. Anti-cancer

drugs targeting this gene must be designed to inhibit both *ACVRL1* and formation of its variants.

ADIPOQ: This gene is specifically expressed in adipose tissue, where it encodes for a protein known as Adinopectin. most common function of the protein is associated with fat metabolism in tissues but apart from that, ADIPOQ has been frequently linked with AMPK, TNF-alpha, NF-kappa-B, and cAMP pathway. Through these cellular transduction and biological pathways, *ADIPOQ* gene can regulate angiogenesis, cellular growth (Bateman et al., 2022), and can restrict uncontrolled cell proliferation (antiangiogenic) by initiating programmed cell death (apoptosis) (Dalamaga et al., 2012). Hence, drug develops must focus on increased activation of *ADIPOQ* genes in breast cancer patients when selecting it as a probable target.

According to Table 5.7, *ADIPOQ* gene has a deleterious variant, G54V, with a highly negative svm score of -3.74. The variant is much likely to code for a protein that can increase the occurrence of cellular mutation and breast cancer proliferation and spreading. It is therefore imperative that drugs designed to target *ADIPOQ* gene can inhibit its variant formation, while enhancing ADIPOQ expression.

TGFBR2: Transforming Growth Factor Beta Receptor 2 (*TGFBR2*) gene encodes a transmembrane protein which, via formation of a complex, coordinates the transcription of genes that can control tumor formation, proliferation and differentiation (Bateman et al., 2022). This genre is part of TGFβ signaling pathway which is commonly responsible for regulation of cellular transition, migration and even cell death (apoptosis). TGFBR2 is often involved in processes relating to cell-cycle disruption and hence it can be concluded that its expression can actively contribute to destruction of cancerous cells and downregulate tumor

growth. Loss of activity of *TGFBR2* gene can promote tumorigenesis due to decreased tumor suppressive activity. In context of breast cancer drug development, *TGFBR2* is an efficient target (Lobato, 2014).

Table 5.7 shows the two deleterious variants of *TGFBR2*, L308P and L333P, with negative svm scores of -2.79 and -2.84 respectively. Successful expression of these variants can give rise to harmful mutations in cell and worsen the condition of breast cancer patients. To prevent that, drugs need to be designed such that snp variant formation of *TGFBR2* gene is inhibited by it.

5.7 Survival Analysis of Genes with potential SNPs

The survival analysis profile was obtained for each gene in the form of a Kaplan-Meier Plot. The plots illustrate the survival probability of patients suffereing from each sub-type of breast cancer when the particular gene is expressed at high or medium-low level. The p-value indicates the difference between multiple survival curves in a KM-plot. The lower the p-value (preferably less than 0.05), the greater the chances of rejecting the null hypothesis that no difference exists between the survival curves.

The figures below illustrate that *TEK, ENG, KDR, NOS3, IGF1, PPARG, ACVRL1, ADIPOQ, TGFBR2* is highly expressed in all 3 major cancer sub-types but its survival probability in each sub-type varies. The p-value points out if this variation is significant or not.

***TEK* (TEK receptor tyrosine kinase) Gene:**

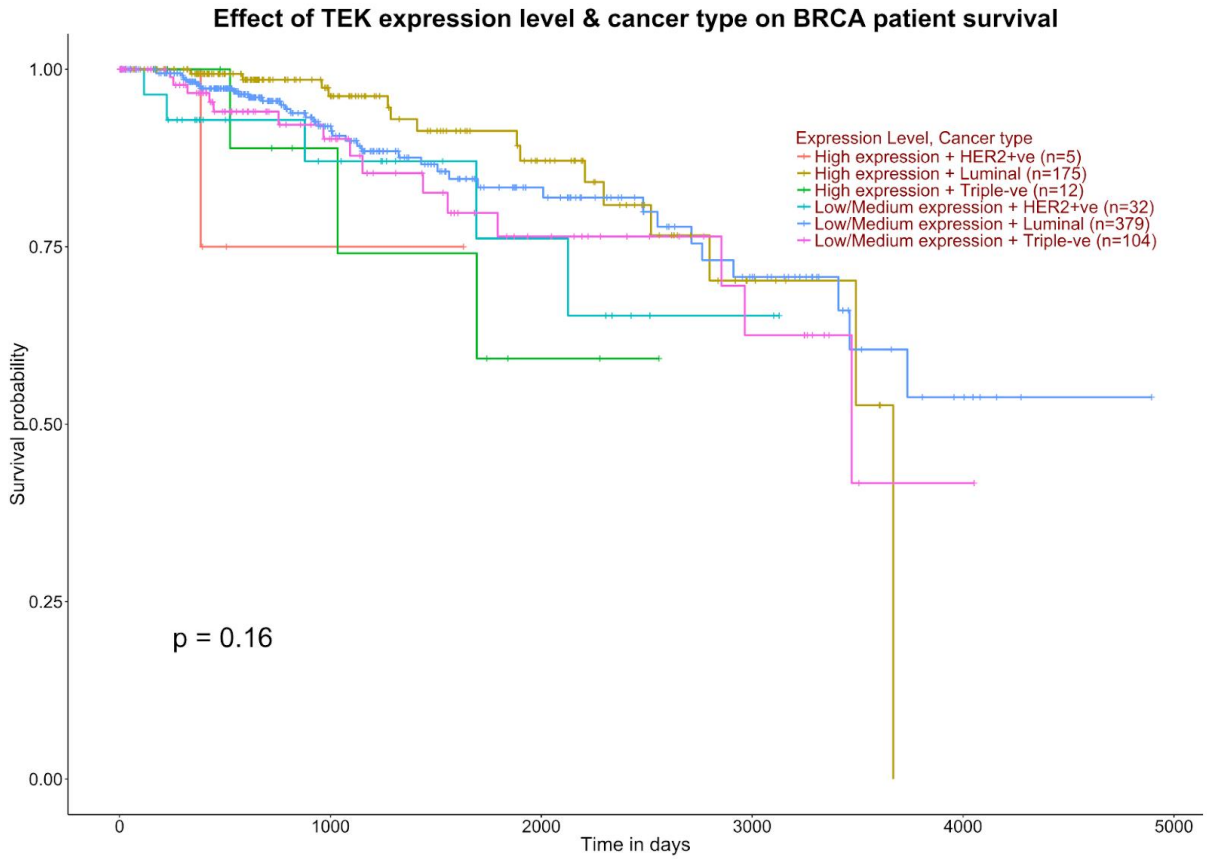

Figure 5.5: Effect of TEK expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 75%, 98% and 100% respectively, following high expression of *TEK* gene. At around 3600 days, high expression of TEK in Luminal breast cancer patients reduces their survival chances to 60% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 40%. It can be understood that TEK expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

***ENG* (endoglin) gene:**

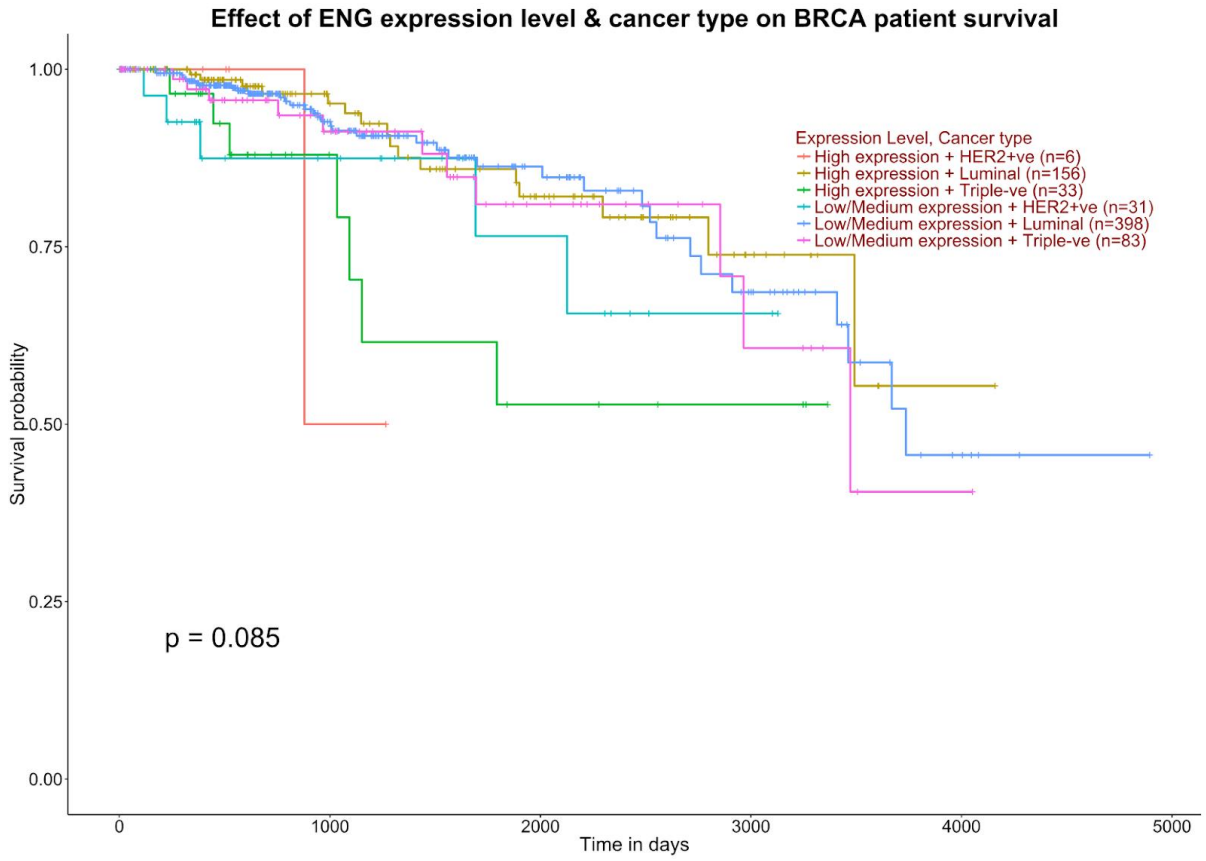

Figure 5.6: Effect of ENG expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 100%, 98% and 90% respectively, following high expression of *ENG* gene. At around 3500 days, medium expression of ENG in Luminal breast cancer patients reduces their survival chances to 60% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 40%. It can be understood that ENG expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

***KDR* (Kinase insert domain receptor) Gene:**

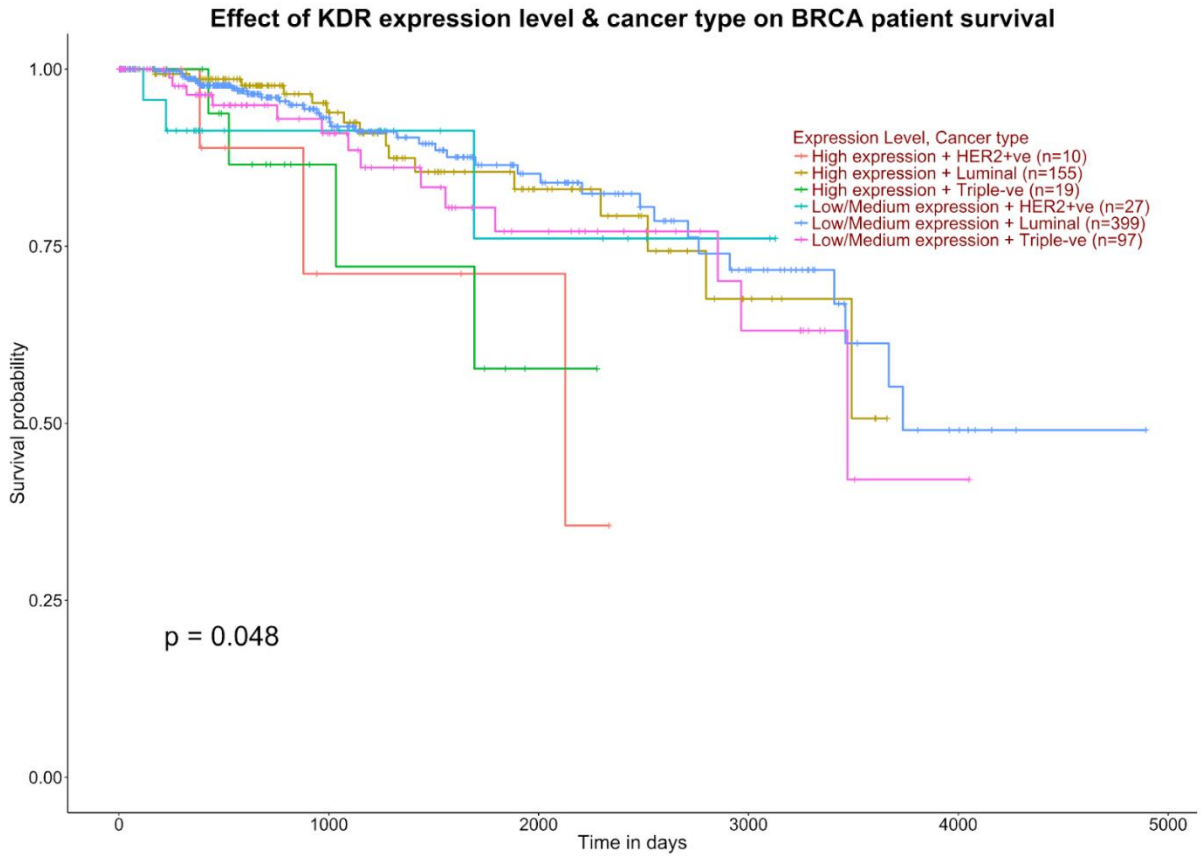

Figure 5.7: Effect of KDR expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 90%, 98% and 87% respectively, following high expression of *KDR* gene. At around 3500 days, medium expression of KDR in Luminal breast cancer patients reduces their survival chances to 65% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 40%. It can be understood that KDR expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

***NOS3* (Nitric Oxide Synthase 3) Gene:**

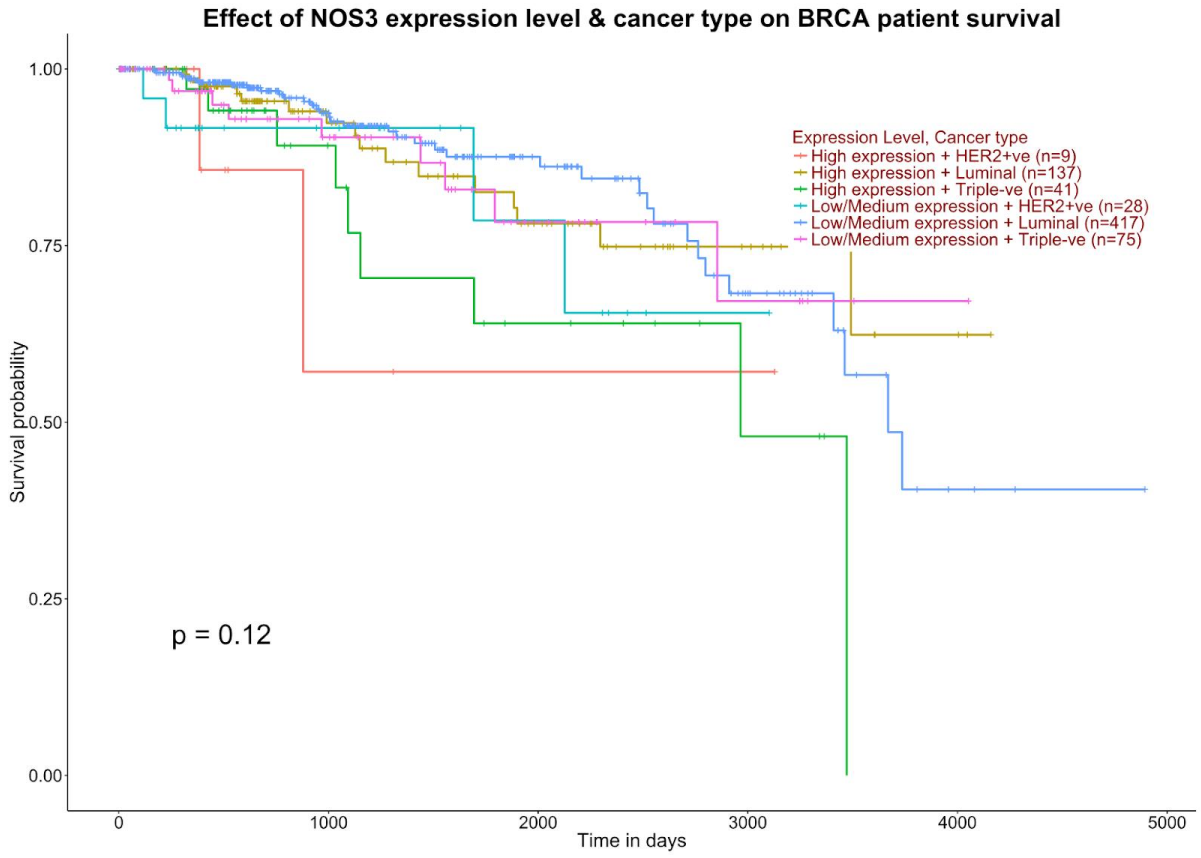

Figure 5.8: Effect of NOS3 expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 84%, 97% and 94% respectively, following high expression of *NOS3* gene. At around 3500 days, high expression of NOS3 gene can reduce survival chances of TNBC patients to almost 0%. Also, medium expression of NOS3 in Luminal breast cancer and Triple negative breast cancer (TNBC) patients reduces patient survival chances to 70%. It can be understood that KDR expression has most significant effect on reducing survival probability of patients suffering from Triple negative breast cancer.

IGF1 (Insulin Like Growth Factor 1) Gene:

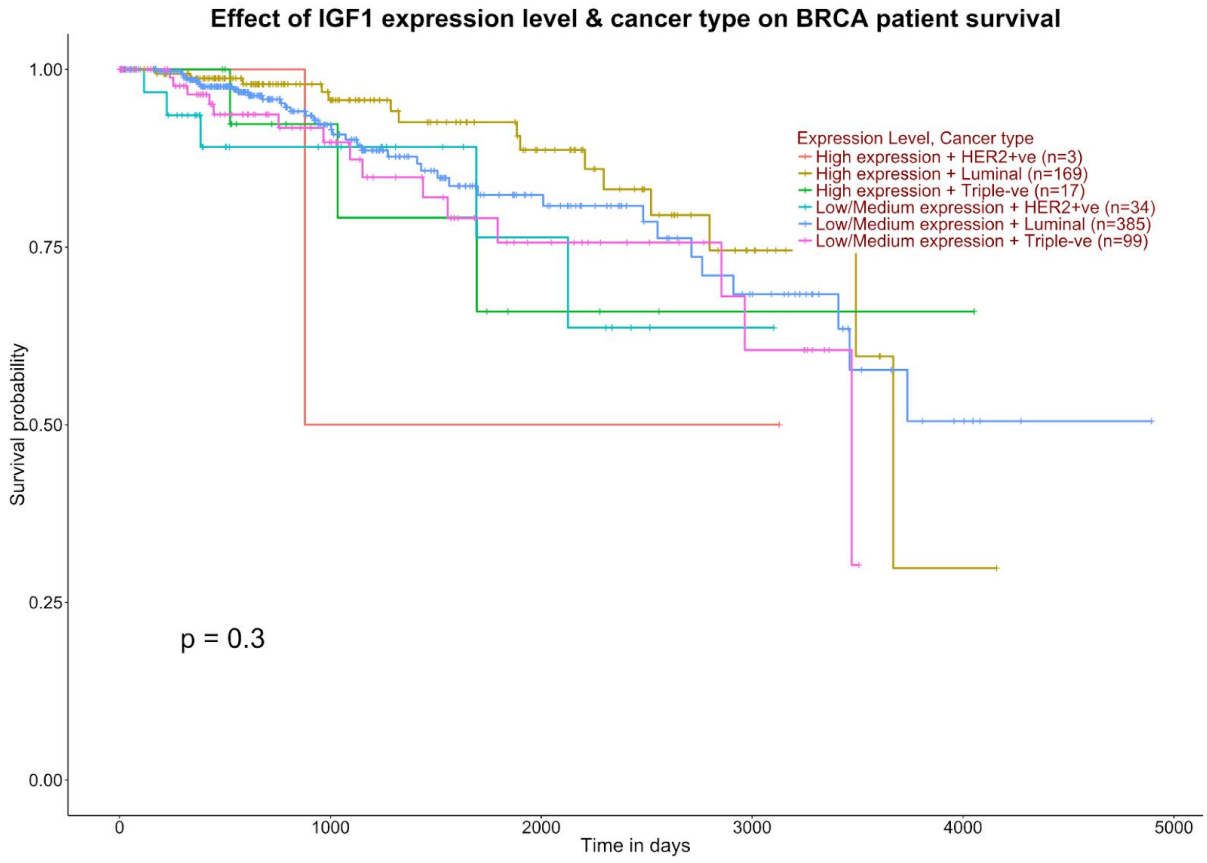

Figure 5.9: Effect of IGF1 expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 100%, 98% and 94% respectively, following high expression of *IGF1* gene. Also, around 2000 days, high expression of IGF1 gene can reduce survival chances of HER2+ve patients to almost 50%. At around 3500 days, medium expression of IGF1 in Luminal breast cancer patients reduces their survival chances to 55% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 30%. It can be understood that IGF1 expression has most significant effect on reducing survival probability of patients suffering from HER2+ve and Triple negative breast cancer.

***PPARG* (Peroxisome Proliferator Activated Receptor Gamma) Gene:**

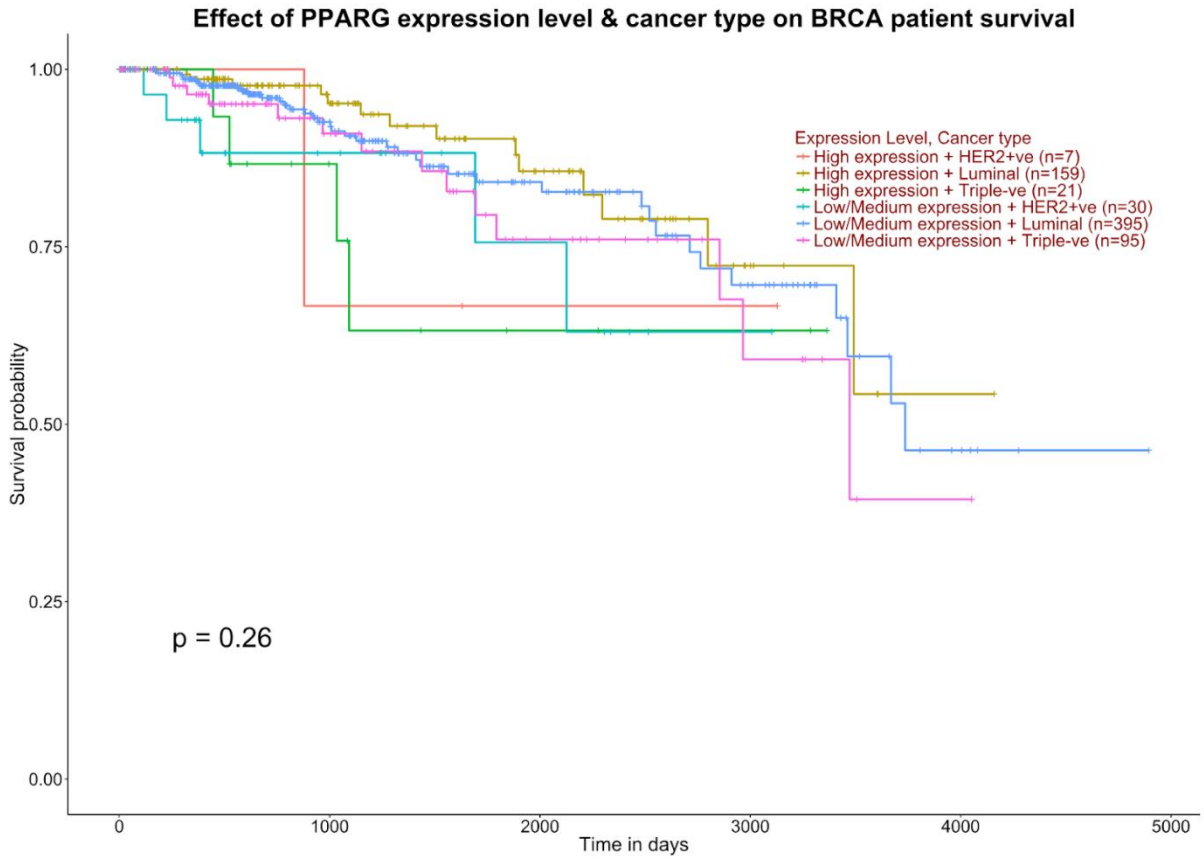

Figure 5.10: Effect of PPARG expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 100%, 98% and 85% respectively, following high expression of *PPARG* gene. Also, around 2000 days, high expression of PPARG gene can reduce survival chances of HER2+ve and TNBC patients to almost 63% and 75% respectively. At around 3500 days, medium expression of PPARG in Luminal breast cancer patients reduces their survival chances to 63% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 40%. It can be understood that PPARG expression has most significant effect on reducing survival probability of patients suffering from Triple negative breast cancer.

***ACVRL1* (Activin A receptor like type 1) Gene:**

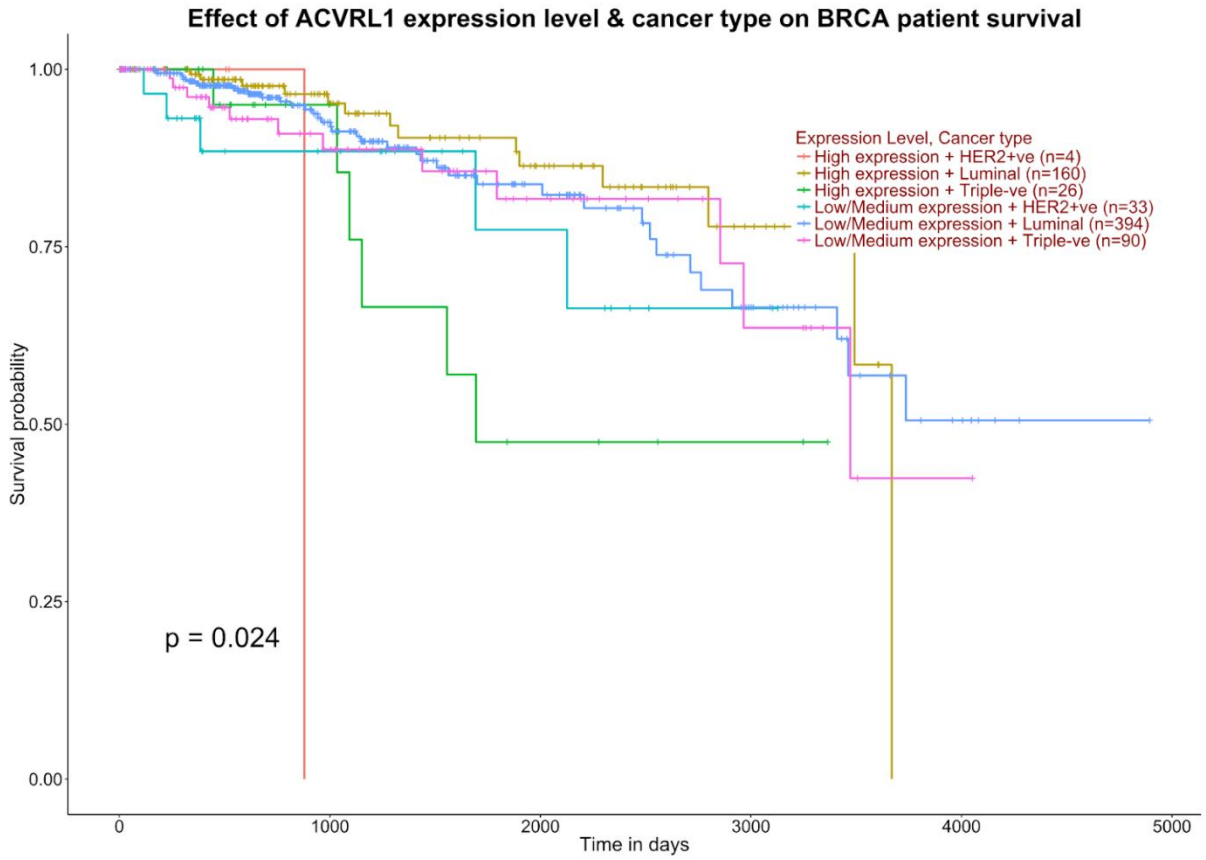

Figure 5.11: Effect of ACVRL1 expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 100%, 98% and 95% respectively, following high expression of *ACVRL1* gene. At around 3500 days, medium expression of ACVRL1 in Luminal breast cancer patients reduces their survival chances to 55% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 40%. It can be understood that ACVRL1 expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

***ADIPOQ* (Adiponectin, C1Q And Collagen Domain Containing) Gene:**

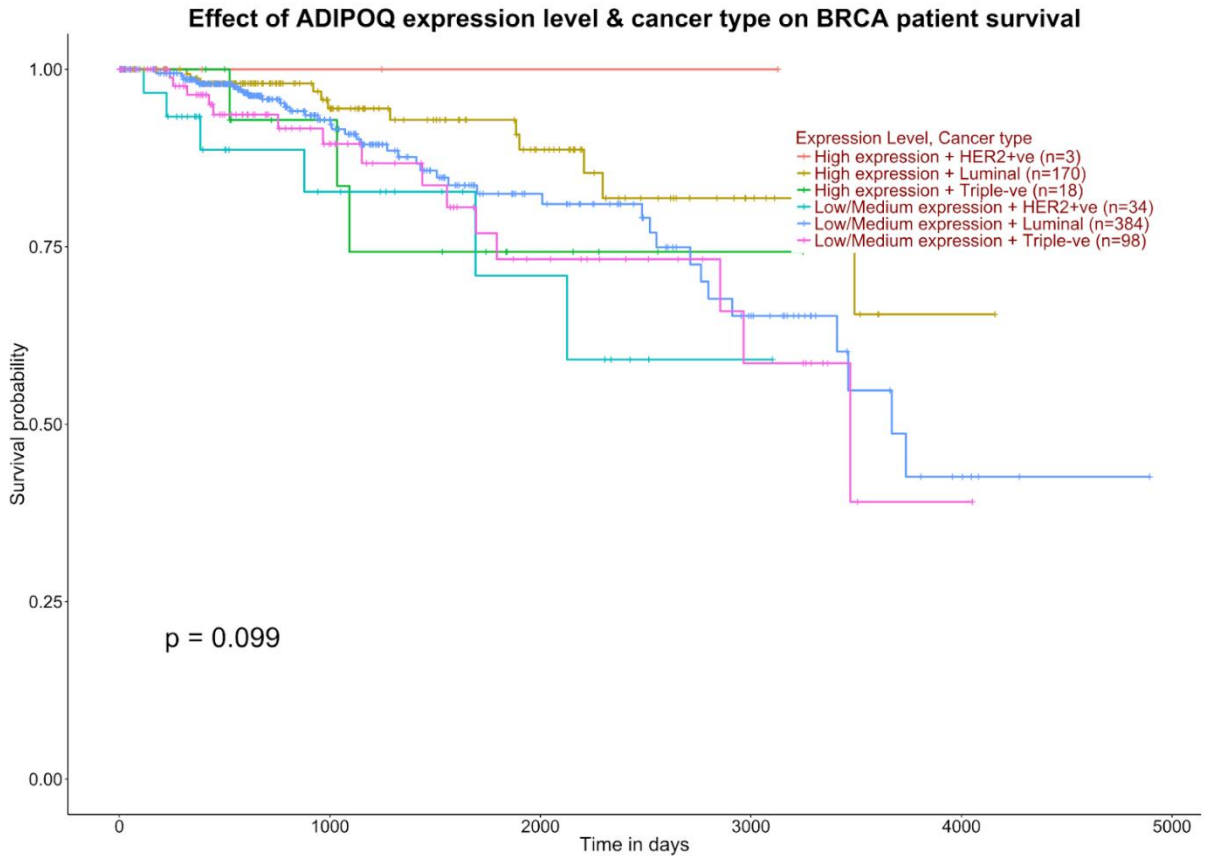

Figure 5.12: Effect of ADIPOQ expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 100%, 98% and 90% respectively, following high expression of *ADIPOQ* gene. At around 3500 days, medium expression of ADIPOQ in Luminal breast cancer patients reduces their survival chances to 54% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 39%. It can be understood that ADIPOQ expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

***TGFBR2* (Transforming Growth Factor Beta Receptor 2) Gene:**

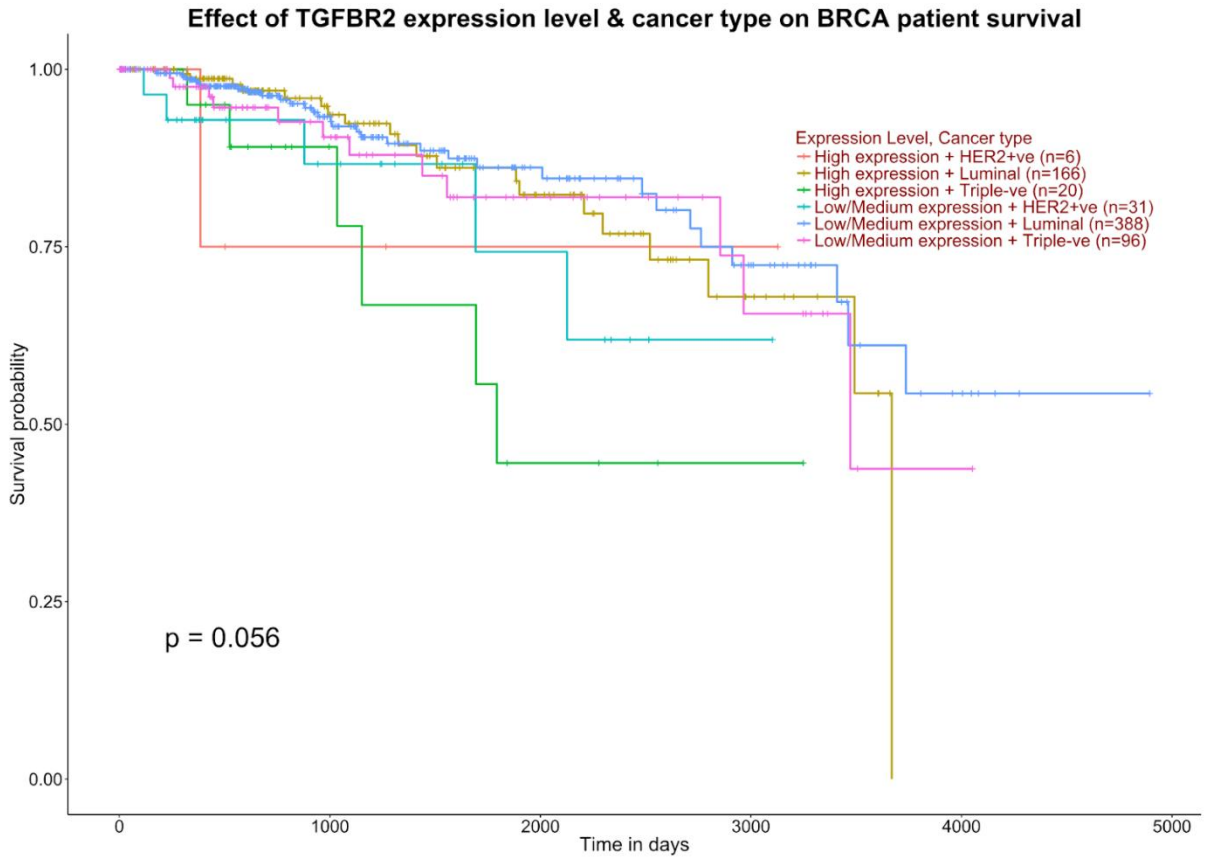

Figure 5.13: Effect of TGFBR2 expression level and cancer type on BRCA patient survival (*UALCAN*, 2022)

For the first 500 days after cancer development, the patients of HER2+ve, Luminal and Triple negative breast cancer have survival chance of 75%, 98% and 88% respectively, following high expression of *TGFBR2* gene. At around 3500 days, medium expression of TGFBR2 in Luminal breast cancer patients reduces their survival chances to 63% whereas medium expression in Triple negative breast cancer (TNBC) patients reduces their survival chances to 42%. It can be understood that TGFBR2 expression has most significant effect on reducing survival probability of patients suffering from Luminal and Triple negative breast cancer.

Chapter 6

Summary

The table below summarizes the contents of Topic 5.6 and Topic 5.7.

Table 5.8: Effect of Gene expression on survival of breast cancer patients.

Name of Gene	p-value	Breast Cancer Type	Survival at High Expression (2000 days)	Survival at Low Expression (2000 days)	Deleterious SNP(s)
TEK	0.16	HER2+ve	75% (extrapolated)	78%	T391I
		Luminal	87%	83%	
		Triple-ve	60%	78%	
ENG	0.085	HER2+ve	50% (extrapolated)	77%	T295R
		Luminal	82%	85%	
		Triple-ve	53%	80%	
KDR	0.048	HER2+ve	72%	75%	V848E, R787G
		Luminal	83%	85%	
		Triple-ve	63%	76%	
NOS3	0.12	HER2+ve	57%	78%	R372C
		Luminal	78%	82%	
		Triple-ve	63%	78%	
IGF1	0.3	HER2+ve	53%	63%	-
		Luminal	83%	75%	
		Triple-ve	65%	82%	
PPARG	0.26	HER2+ve	65%	75%	R316H
		Luminal	85%	82%	
		Triple-ve	63%	75%	
ACVRL1	0.024	HER2+ve	-	77%	R374W, I245N, G211D
		Luminal	85%	80%	
		Triple-ve	47%	80%	

Table 5.8 (continued)

Name of Gene	p-value	Breast Cancer Type	Survival at High Expression (2000 days)	Survival at Low Expression (2000 days)	Deleterious SNP(s)
ADIPOQ	0.099	HER2+ve	100%	77%	G54V
		Luminal	83%	81%	
		Triple-ve	75%	74%	
TGFBR2	0.056	HER2+ve	75%	75%	L333P, L308P
		Luminal	85%	82%	
		Triple-ve	45%	80%	

Most genes in the table, namely: TEK, ENG, NOS3, IGF1, PPARG, ADIPOQ, TGFBR2, have a p-value greater than 0.05, which indicates that the survival curves for each breast cancer type do not differ significantly. The main objective was to tabulate the consequence of gene expression on the survival percentage of patients. It can therefore be understood that in case of all genes, their expression expression most significantly reduces the survival chances of patients suffering from Triple negative breast cancer (TNBC). The table also mentions the important deleterious variant(s) (due to single nucleotide polymorphism) of each gene which, when expressed, can induce mutation and tumor progression.

It is evident from the table that TNBC patients have higher chances of succumbing to breast cancer. This makes it a matter of consideration for drug developers, who must acknowledge the necessity of developing a drug that can effectively fight this sub-type of breast cancer.

Chapter 7

Conclusion

The NOTCH pathway is an extremely resourceful pathway that is involved in the regulation of genes associated with a vast range of biological process. This makes it a promising target for anti-cancer drug development. Data extracted and interpreted so far helped us narrow down a number of genes: *TEK, ENG, KDR, NOS3, IGF1, PPARG, ACVRL1, ADIPOQ, TGFBR2* which can exert significant effect on the survival of certain of breast cancer patients when their expression is up-regulated or down-regulated via external chemical stimulus. It can be interpreted that these genes and their expression has most effect on Triple-negative breast cancer, which has the worst prognosis among breast cancer sub-types and can only be managed via chemotherapy as no known drugs have been developed yet. Drug development focused on the inhibition or negative regulation of these genes and their expressed protein may finally be able to carve a way to prolong and cure the survival of breast cancer patients, particularly TNBC.

The 9 selected genes can be referred to as a “panel of genes” that can be used as diagnostic markers whose level of expression (after anti-cancer drug therapy) can be used to analyse cancer stage and prognosis.

Chapter 8

Significance of the study

For decades, the NOTCH pathway has been regarded as an important genes and some of its associated genes have been researched extensively. However, there is a gap in those works as it has not managed to focus on all NOTCH associated genes and their ability to regulate breast cancer cell proliferation. This paper, not only fills that void but also delves into how NOTCH co-expressed genes can directly and indirectly effect other pathways of human breast cancers. In addition to that, the data and results compiled in this paper can be used as a rich source of information for future studies as well.

Advancements in technology and research has allowed us to expand the perimeter of treatment options beyond drugs and oral medication. The implementation of appropriate gene therapy techniques in cancer management has been in the clinical trial stage for a long time and this paper was constructed with the hopes of contributing to it. Gene therapy involves the administration of genetic material to manipulate a gene's expression or modify the general characteristics of targeted cells. Some commonly employed gene therapy methods are: gene inhibition therapy, gene augmentation therapy, suicidal gene therapy etc. Appropriate gene therapy method, accompanied by a suitable vector can be used to inhibit dangerous genes and promote desirable genes and eventually mend the function of tumor cells before harmful proteins are even expressed. Oncotherapy involving engineered virus can be employed to trigger destruction of cancer cells.

Apart from compiling a short list of 9 favorable genes from more than 3500 breast cancer genes, this paper also provides an insight into some of the essential snp variants of each gene. The presence of deleterious and non-deleterious variants is a vital factor to acknowledge and examine in order to develop a drug that can successfully regulate the gene and its variants to exert the best anti-cancer effect possible. Any therapeutic option targeting these gene must be

able to inhibit the deleterious variant's expression and promote the non-deleterious variant's expression.

The utilization of SMIs (small molecule inhibitors) is a convenient method of inhibiting the protein expressed by the target gene. Its notably small size allows it to access and suppress its targets easily. Oral drugs designed and produced in the form of SMIs can be used to selectively inhibit the proteins expressed by the deleterious snp variants of the genes under investigation.

The role of monoclonal antibodies (mAbs) as an efficient therapeutic agent must be reckoned in this case. Monoclonal antibodies can be designed such that its is able to recognize antigens on proteins expressed by unfavorable genes or variants with negative svm score (deleterious) or even capable of inducing an immune response against antigens present on cells where tumor-proliferating genes are overexpressed.

A new and propitious approach to cancer therapy is the concept of mRNA vaccines. The success of mRNA-based COVID vaccines has encouraged researchers to focus on its use a cancer treatment option. A mRNA cancer vaccine contains mRNA which, once inside a healthy cell, will instruct the cell to produce an antigen or market substance generally found on surface of cancer cells. The immune system is subsequently activated into producing antibodies and fighting the supposed cancer cells. This enables the body to create a faster defense against the proliferation, growth and invasion of cancer cells, which are killed or neutralized just as soon their presence is detected.

This paper is the first to focus on the 9 potential NOTCH-correlated breast cancer genes whose expression level can be utilized to examine the survival of breast cancer patients, determine the efficacy of the anti-cancer drug and evaluate the prognosis of chemotherapy in cancer patient group at different stages of breast cancer. The paper proposes a panel of 9 genes that can help assess the relation between the expression level of the gene's protein and

the survival and aggressive nature of the specific type of breast cancer. Immunohistochemical Staining can be used to investigate the level of expression of these genes and their respective proteins in clinical samples collected from Bangladesh's cancer population. Furthermore, it can be established if the panel of genes can be used to evaluate the prognosis of breast cancer.

References

Bateman, A., Martin, M.-J., Orchard, S., Magrane, M., Ahmad, S., Alpi, E., Bowler-Barnett, E. H., Britto, R., Hema Bye-A-Jee, Austra Cukura, Denny, P., Dogan, T., Ebenezer, T., Fan, J., Garmiri, P., Jose, L., Hatton-Ellis, E., Hussein, A., Alexandr Ignatchenko, & Insana, G. (2022). UniProt: the Universal Protein Knowledgebase in 2023. Nucleic Acids Research, 51(D1), D523–D531. https://doi.org/10.1093/nar/gkac1052

Bayat, A. (2002). Science, medicine, and the future: Bioinformatics. The BMJ, 324(7344), 1018–1022. https://doi.org/10.1136/bmj.324.7344.1018

BeLow, M., & Clodia Osipo. (2020). Notch Signaling in Breast Cancer: A Role in Drug Resistance. Cells, 9(10), 2204–2204. https://doi.org/10.3390/cells9102204

Cancer today. (2020). Iarc.fr. https://gco.iarc.fr/today/home

CDCBreastCancer. (2023, July 27). Triple-Negative Breast Cancer. Centers for Disease Control and Prevention. https://www.cdc.gov/cancer/breast/triple-negative.htm

Chicco, D., & Agapito, G. (2022). Nine quick tips for pathway enrichment analysis. PLOS Computational Biology, 18(8), e1010348–e1010348. https://doi.org/10.1371/journal.pcbi.1010348

Citra Dewi, Adryan Fristiohady, Riezki Amalia, Nur, Ibrahim, S., & Muchtaridi Muchtaridi. (2022). Signaling Pathways and Natural Compounds in Triple-Negative Breast Cancer Cell Line. Molecules, 27(12), 3661–3661. https://doi.org/10.3390/molecules27123661

Dalamaga, M., Diakopoulos, K. N., & Mantzoros, C. S. (2012). The Role of Adiponectin in Cancer: A Review of Current Evidence. Endocrine Reviews, 33(4), 547–594. https://doi.org/10.1210/er.2011-1015

Erichsen, H. C., & Chanock, S. J. (2004). SNPs in cancer research and treatment. British Journal of Cancer, 90(4), 747–751. https://doi.org/10.1038/sj.bjc.6601574

Fouad, Y. A., & Aanei, C. (2017). Revisiting the hallmarks of cancer. American Journal of Cancer Research, 7(5), 1016–1036. https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5446472/

GeneMANIA. (2024). Genemania.org. https://genemania.org/search/homo-sapiens/SFRP1/TEK/CAV1/ENG/PPARG/ACVRL1/AKT3/EDN1/FLT4/KDR/NOS3/TGFBR2/GATA3/IGF1/CXCL12/AMOTL1/ADAMTS1/DUSP1/ADIPOQ

GEPIA 2. (2024). Cancer-Pku.cn. http://gepia2.cancer-pku.cn/#degenes

GLOBOCAN 2020: New Global Cancer Data. (2020). UICC. https://www.uicc.org/news/globocan-2020-new-global-cancer-data

Hanahan, D., & Weinberg, R. A. (2000). The Hallmarks of Cancer. Cell, 100(1), 57–70. https://doi.org/10.1016/s0092-8674(00)81683-9

Hanahan, D., & Weinberg, R. A. (2011). Hallmarks of Cancer: The Next Generation. Cell, 144(5), 646–674. https://doi.org/10.1016/j.cell.2011.02.013

Hanahan, D., & Weinberg, R. A. (2017). Biological Hallmarks of Cancer. Holland-Frei Cancer Medicine, 1–10. https://doi.org/10.1002/9781119000822.hfcm002

Huber, M. A., Ninel Azoitei, Baumann, B., Grünert, S., Sommer, A., Pehamberger, H., Kraut, N., Hartmut Beug, & Wirth, T. (2004). NF-κB is essential for epithelial-mesenchymal transition and metastasis in a model of breast cancer progression. Journal of Clinical Investigation, 114(4), 569–581. https://doi.org/10.1172/jci21358

Li, D.-H., Liu, X.-K., Tian, X.-T., Liu, F., Yao, X.-J., & Dong, J.-F. (2023). PPARG: A Promising Therapeutic Target in Breast Cancer and Regulation by Natural Drugs. Ppar Research, 2023, 1–18. https://doi.org/10.1155/2023/4481354

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? Trends in Biochemical Sciences, 41(3), 211–218. https://doi.org/10.1016/j.tibs.2015.12.001

Liberti, M. V., & Locasale, J. W. (2016). The Warburg Effect: How Does it Benefit Cancer Cells? Trends in Biochemical Sciences, 41(3), 211–218. https://doi.org/10.1016/j.tibs.2015.12.001

Lobato, F. (2014). ACVRL1 (activin A receptor type II-like 1). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/569/acvrl1-(activin-a-receptor-type-ii-like-1)

Lobato, F. (2014). TGFBR2 (Transforming Growth Factor, Beta Receptor II (70/80kDa)). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/372/tgfbr2-(transforming-growth-factor-beta-receptor-ii-(70-80kda))

Lobato, F. (2015). KDR (kinase insert domain receptor)/Vascular Endothelial Growth Factor Receptor 2 (VEGFR2). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/41055/kdr-(kinase-insert-domain-receptor)-vascular-endothelial-growth-factor-receptor-2-(vegfr2)

Lobato, F. (2022). ENG (endoglin). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/40452/eng-(endoglin)

Lobato, F. (2024). IGF1 (Insulin-Like Growth Factor 1 (Somatomedin C)). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/40927/igf1-(insulin-like-growth-factor-1-(somatomedin-c))

Lobato, F. (2024). TEK (TEK tyrosine kinase, endothelial). Atlasgeneticsoncology.org. https://atlasgeneticsoncology.org/gene/42517/tek-(tek-tyrosine-kinase-endothelial)

Majumder, S., Crabtree, J. S., Golde, T. E., Minter, L. M., Osborne, B. A., & Miele, L. (2020). Targeting Notch in oncology: the path forward. Nature Reviews Drug Discovery, 20(2), 125–144. https://doi.org/10.1038/s41573-020-00091-3

Nandi, A., & Chakrabarti, R. (2020). The many facets of Notch signaling in breast cancer: toward overcoming therapeutic resistance. Genes & Development, 34(21-22), 1422–1438. https://doi.org/10.1101/gad.342287.120

Notch Signaling Pathway | Cell Signaling Technology. (2014). Cell Signaling Technology. https://www.cellsignal.com/pathways/notch-signaling-pathway

Ossovskaya, V. (2016). Exploring Molecular Pathways of Triple-Negative Breast Cancer - Valeria Ossovskaya, Yipeng Wang, Adam Budoff, Qiang Xu, Alexander Lituev, Olga Potapova, Gordon Vansant, Joseph Monforte, Nikolai Daraselia, 2011. Genes & Cancer. https://journals.sagepub.com/doi/full/10.1177/1947601911432496

Panagiotis Ntziachristos, Jing Shan Lim, Sage, J., & Iannis Aifantis. (2014). From Fly Wings to Targeted Cancer Therapies: A Centennial for Notch Signaling. Cancer Cell, 25(3), 318–334. https://doi.org/10.1016/j.ccr.2014.02.018

Polyak, K., & Weinberg, R. A. (2009). Transitions between epithelial and mesenchymal states: acquisition of malignant and stem cell traits. Nature Reviews Cancer, 9(4), 265–273. https://doi.org/10.1038/nrc2620

Previs, R. A., Coleman, R. L., Harris, A. L., & Sood, A. K. (2015). Molecular Pathways: Translational and Therapeutic Implications of the Notch Signaling Pathway in Cancer. Clinical Cancer Research, 21(5), 955–961. https://doi.org/10.1158/1078-0432.ccr-14-0809

Rojas, K., & Stuckey, A. (2016). Breast Cancer Epidemiology and Risk Factors. Clinical Obstetrics and Gynecology, 59(4), 651–672. https://doi.org/10.1097/grf.0000000000000239

Sereno. (2021, March 14). Comparison of Pearson vs Spearman Correlation Coefficients. Analytics Vidhya. https://www.analyticsvidhya.com/blog/2021/03/comparison-of-pearson-and-spearman-correlation-coefficients/#:~:text=Pearson%20and%20Spearman%20correlation%20coefficients,coefficient%20evaluates%20the%20monotonic%20relationship

Shao, S., Zhao, X., Zhang, X., Luo, M., Zuo, X., Huang, S., Wang, Y., Gu, S., & Zhao, X. (2015). Notch1 signaling regulates the epithelial–mesenchymal transition and invasion of breast cancer in a Slug-dependent manner. Molecular Cancer, 14(1). https://doi.org/10.1186/s12943-015-0295-3

Shastry, B. S. (2007). SNPs in disease gene mapping, medicinal drug development and evolution. Journal of Human Genetics, 52(11), 871–880. https://doi.org/10.1007/s10038-007-0200-z

ShinyGO v0.741: Gene Ontology Enrichment Analysis + more. (2022). Sdstate.edu. http://bioinformatics.sdstate.edu/go74/

Signal Transduction. (2023). Reactome.org. https://reactome.org/content/detail/R-HSA-162582

Single Nucleotide Polymorphism (SNP) Allele Frequency DNA Pools | Society for Mucosal Immunology. (2024). Socmucimm.org. https://www.socmucimm.org/resources/news-media/single-nucleotide-polymorphism-snp-allele-frequency-dna-pools/

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=ACVRL1&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=ADIPOQ&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=ENG&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=IGF1&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=KDR&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=NOS3&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=PPARG&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=TEK&ctype=BRCA

UALCAN. (2022). Uab.edu. https://ualcan.path.uab.edu/cgi-bin/TCGA-survival1.pl?genenam=TGFBR2&ctype=BRCA

UniProt. (2024). Uniprot.org. https://www.uniprot.org/locations/SL-9905

Viatour, P., Ehmer, U., Saddic, L. A., Dorrell, C., Andersen, J. B., Lin, C., Anne Flore Zmoos, Mazur, P. K., Schaffer, B. E., Ostermeier, A., Vogel, H., Sylvester, K. G., Thorgeirsson, S. S., Grompe, M., & Sage, J. (2011). Notch signaling inhibits hepatocellular carcinoma following inactivation of the RB pathway. Journal of Experimental Medicine, 208(10), 1963–1976. https://doi.org/10.1084/jem.20110198

Winters, S., Martin, C., Murphy, D., & Shokar, N. K. (2017). Breast Cancer Epidemiology, Prevention, and Screening. Progress in Molecular Biology and Translational Science, 1–32. https://doi.org/10.1016/bs.pmbts.2017.07.002

Wolfe, M. S. (2020). Unraveling the complexity of γ-secretase. Seminars in Cell & Developmental Biology, 105, 3–11. https://doi.org/10.1016/j.semcdb.2020.01.005

Yang, F., Zhou, X., Miao, X., Zhang, T., Hang, X., Tie, R., Liu, N., Tian, F., Wang, F., & Yuan, J. (2014). MAGEC2, an epithelial-mesenchymal transition inducer, is associated with breast cancer metastasis. Breast Cancer Research and Treatment, 145(1), 23–32. https://doi.org/10.1007/s10549-014-2915-9

Yi Sheng Sun, Zhao, Z., Zhang Nv Yang, Xu, F., Hang Jing Lu, Zhi Yong Zhu, Shi, W., Jiang, J., Ping Ping Yao, & Han Ping Zhu. (2017). Risk Factors and Preventions of Breast Cancer. International Journal of Biological Sciences, 13(11), 1387–1397. https://doi.org/10.7150/ijbs.21635

Zhong, Z., Yu, J., Virshup, D. M., & Madan, B. (2020). Wnts and the hallmarks of cancer. Cancer and Metastasis Reviews, 39(3), 625–645. https://doi.org/10.1007/s10555-020-09887-6

Zhu, C., Baumgarten, N., Wu, M., Wang, Y., Arka Provo Das, Kaur, J., Fatemeh Behjati Ardakani, Thanh Thuy Duong, Minh Duc Pham, Duda, M., Dimmeler, S., Yuan, T., Schulz, M. H., & Krishnan, J. (2023). CVD-associated SNPs with regulatory potential reveal novel non-coding disease genes. Human Genomics, 17(1). https://doi.org/10.1186/s40246-023-00513-4

Zou, D., Li, Z., Fei Lv, Yang, Y., Yang, C., Song, J., Chen, Y., Jin, Z., Zhou, J., Jiang, Y., Ma, Y., Tang, Y., & Zhang, Y. (2021). Pan-Cancer Analysis of NOS3 Identifies Its Expression and Clinical Relevance in Gastric Cancer. Frontiers in Oncology, 11. https://doi.org/10.3389/fonc.2021.592761