CLINICOPATHOLOGICAL AND GENOMIC MUTATIONAL ANALYSIS OF TRIPLE NEGATIVE BREAST CANCER IN MIZO POPULATION

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

LALAWMPUII PACHUAU

MZU REGN NO.: 1807391

Ph.D. REGN NO.: MZU/Ph.D./1292 of 07.08.2018

DEPARTMENT OF BIOTECHNOLOGY

SCHOOL OF LIFE SCIENCES

NOVEMBER 2023

CLINICOPATHOLOGICAL AND GENOMIC MUTATIONAL ANALYSIS OF TRIPLE NEGATIVE BREAST CANCER IN MIZO POPULATION

BY

LALAWMPUII PACHUAU

Department of Biotechnology

Supervisor: Prof. N. SENTHIL KUMAR

Submitted

In partial fulfillment of the requirement of the degree of Doctor of Philosophy in Biotechnology of Mizoram University, Aizawl

MIZORAM UNIVERSITY

(A Central University)

Department of Biotechnology

Tanhril, Aizawl - 796004

Gram: MZU, Phone: 9436352574, Email: nskmzu@gmail.com

CERTIFICATE

This is to certify that the thesis entitled **"Clinicopathological and Genomic Mutational Analysis of Triple Negative Breast Cancer in Mizo Population"** submitted by **Lalawmpuii Pachuau**, Ph.D. Scholar for the award of the Degree of Doctor of Philosophy in Biotechnology is carried out under my supervision and incorporates the student bona-fide research and this has not been submitted for the award of any degree in this or any other university or institute of learning.

Place: Aizawl, Mizoram (PROF. N. SENTHIL KUMAR)

Date: Supervisor

MIZORAM UNIVERSITY

Month: November **Year: 2023**

DECLARATION

I **Lalawmpuii Pachuau,** hereby declare that the subject matter of this thesis is the record of work done by me, that the contents of this thesis did not form basis of the award of any previous degree to me or to the best of my knowledge to anybody else, and that the thesis has not been submitted by me for any research degree in any other University/Institute.

This is being submitted to the Mizoram University for the degree of **Doctor of Philosophy in Biotechnology.**

(**LALAWMPUII PACHUAU**)
Research Scholar

(PROF. JOHN ZOTHANZAMA)
Head of Department

(PROF. N. SENTHIL KUMAR)
Supervisor

ACKNOWLEDGEMENT

I could not have undertaken this journey without my guide The Prof N. Senthil Kumar, an esteemed Professor in the Department of Biotechnology at Mizoram University who has given me the chance, ever encouraging me, assisting me in every step with his immense knowledge and expertise. Thank you for giving me this opportunity and believing in me.

My sincere appreciation to Prof. John Zothanzama, the Head of the Department of Biotechnology. And equally indebted to Dr.H.Lalhruaitluanga, Assistant Professor in the department.

This work would not have seen the light of day if not for Remmawia, Research Associate who lends me his vast knowledge in the field of genetics, decoding the data and make it decipherable when the tough gets going, in the nick of time.

Tei, Technical Assistance for your immense contribution from DNA to data, I could never have thank you enough, you were always just a phone call away. You are certainly a Blessing, but not in disguise.

Zami, my lab mate, thank you for showing and leading me in the intricate helix and spiral coil during research days. Youth is definitely not wasted on you.

Vignesh K and Alex, for a helping hand and your time whenever needed during my thesis work.

A special thanks to my colleagues Dr.Lalchhandami, Head of Pathology Department and Dr.C.Lalchhandama Pathologist at Civil Hospital Aizawl who never made my absence felt even when my attendance in the Department was triple negative for few consecutive days. Dr. K.L.Remsanga and Dr.John Zohmingthanga, the then Medical Superintendent and Deputy Medical Superintendent Civil Hospital Aizawl for paving the way for me to take my first step in the field of research. Dr.Saia Chenkual, the then Head of Department of Surgery and other surgeons from

Civil Hospital Aizawl for their cooperation in this work.

I am thankful to Baby for collecting the samples, relevant lumps and bumps clinical details and documentation which are a treasure trove.

I am especially grateful to my fellow lab mates, Ranjan, Andrew, Temawii and all the research scholars of the Department of Biotechnology, Mizoram University. We are complimentary to one another, forming a friendship bond over the years.

I am much obliged to the DBT- Advanced State Biotech Hub at Biotechnology Department, MZU and National Institute of Biomedical Genomics (NIBMG), Kalyani for the laboratory facility during the research.

I am Forever grateful to my family for the foundation stone in my education, for adapting well to the Odd working hours that comes with my Profession. My elder brother Mama who would have been proud of me in his own subtle way. To my husband who thought this journey has been a long haul especially for Him alone. You have learnt well that temperament is directly proportional to stress level. To Gabrielle who came into my life midway in this research period, You have made me stronger, better and more fulfilled than I could have ever imagined.

Above all these, there is God, thank you for my life, please take care of my Patients.

Place: Aizawl **(LALAWMPUII PACHUAU)**

Table of Contents

Contents **Page No.**

Contents	Page No.
Certificate	i
Declaration	ii
Acknowledgements	iii-iv
Table of Contents	v
List of Figures	vi
List of Tables	vii-viii
Abbreviations	ix-x
Chapter 1: Introduction	1-4
Chapter 2: Review of Literature	5-15
Chapter 3: Aim and Objectives	16
Chapter 4: Materials and Methods	17-27
Chapter 5: Results	28-101
Chapter 6: Discussion	102-108
Chapter 7: Summary	109-111
Appendices: Appendix I	112-115
Appendix II	116-117
Bibliography	110-127

List of Figures

Figure 1. Macroscopic structure of the breast and changes during lactation.

Figure 2. Hallmarks of Cancer.

Figure 3. Tumor metastasis in Lymph node.

Figure 4. Workflow for Exome Sequencing Library Preparation - Using Illumina Truseq Exome with pre- capture pooling.

Figure 5. IHC slide images of representative samples.

Figure 6. BMI of TNBC patients.

Figure 7. Overall survival of TNBC patients.

Figure 8. Overall survival in TNBC patients based on body weight.

Figure 9. Summary of the filtered somatic variants in TNBC.

Figure 10. Top 10 frequently mutated genes showing protein mutation rate and the sites of variants.

Figure 11. Interacting gene pair analysis of the 30 most mutated genes in somatic TNBC.

Figure 12. Mutational signature analysis with COSMIC database in somatic TNBC.

Figure 13. COSMIC mutational signatures across somatic TNBC samples.

Figure 14. Average signature exposure and mutation load in somatic TNBC samples.

Figure 15. Druggable genes and druggable categories identified in somatic TNBC.

Figure 16. Mutational landscape in germline variants in TNBC.

Figure 17. Interacting gene pair analysis of the 30 most mutated genes in germline TNBC.

Figure 18. Mutational signature in germline TNBC matching with COSMIC signature database.

Figure 19. Druggable genes and druggable categories identified in germline variants in TNBC.

Lists of Tables

Table 1. Frequency analysis of sociodemographic variables among the TNBC patients.

Table 2. Food habits of TNBC patients.

Table 3. Lifestyle habits of TNBC patients.

Table 4. Biological and health status of TNBC patients.

Table 5. Clinical Parameters of TNBC patients.

Table 6. Association between Demographic Variables and food habits Versus the recurrence of TNBC.

Table 7. Association between biological parameters of TNBC patients with recurrence.

Table 8. Association between socio-demographic variables and biological parameters of TNBC patients with other types of cancer.

Table 9. Association between socio demographic variables and biological parameters of TNBC patients with Co-morbidities.

Table 10. Association between socio demographic variables and biological parameters of TNBC patients with Familial_Nonfamilial occurrence.

Table 11. Association between socio demographic variables and biological parameters of TNBC patients with Tumour type.

Table 12. Association between socio demographic variables and biological parameters of TNBC patients with Lymph node occurrence.

Table 13. Association between socio demographic variables and biological parameters of TNBC patients with DCIS occurrence.

Table 14. Association between socio demographic variables and biological parameters of TNBC patients with Survival.

Table 15. Association between socio demographic variables between Healthy controls and TNBC patients.

Table 16. Food habits between Healthy controls and TNBC patients.

Table 17. Life style habits between Healthy controls and TNBC patients.

Table 18. Health related parameters between Healthy controls and TNBC patients.

Table 19. Summary of somatic variants analysis.

Table 20. Summary of somatic variants per samples.

Table 21. Top 10 frequently mutated genes and the number of variants and altered samples.

Table 22. Mutational signature analysis of somatic variants in TNBC.

Table 23. Oncogenic pathways associated with variants and their related somatic genes in TNBC.

Table 24. Potential Druggable genes identified in somatic genes in TNBC.

Table 25. Summary of germline variants analysis.

Table 26. Top 10 frequently mutated genes and the number of Missense mutations and altered samples in germline TNBC.

Table 27. Significantly co-occuring gene pairs and mutually exclusive gene pairs in germline TNBC.

Table 28. Mutational signature genes match with COSMIC database in germline TNBC.

Table 29. Oncogenic pathways associated with variants and their related genes in germline TNBC.

Table 30. Summary of clinical data analysis displaying clinical characteristics and corresponding variant counts in identified genes.

Abbreviations

IBC - Inflammatory BC

LCIS -Lobular carcinoma in situ

DCIS -Ductal carcinoma in situ

IDC - Invasive or Infiltrating Ductal Carcinoma

ILC -Invasive or Infiltrating Lobular Carcinoma

RR - Relative Risk

MSCI - Mizoram State Cancer Institute

BC - Breast Cancer

PCR - Polymerase Chain Reaction

SNPs - Single Nucleotide Polymorphisms

Indels - Insertions and deletions

rCRS - revised Cambridge Reference Sequence

ER/ PR - Hormone Receptors

ER - Estrogen Receptor

PR - Progesterone Receptor

HER2/neu - Human epidermal growth factor receptor 2

APF - Affects Protein Function

T - Tolerated

PD - Possibly Damaging

PRD - Probably Damaging

B - Benign

% - Percentage

mtDNA - Mitochondrial DNA

µl - Micro litre

oC -Degree Celsius

PCR -Polymerase Chain Reaction

mM -Milli Molar

ng -Nano Gram

maf -Minor Allele Frequency

kb -Kilo base

dNTPs -Deoxynucleotide Triphosphate

OR -Odds Ratio

CI -Confidence Interval

Ref -Reference

< -Less Than

> -Greater Than

+ve -Positive

-ve -Negative

LIQ -Lower Inner Quadrant

LOQ -Lower Outer Quadrant

UIQ -Upper Inner Quadrant

UOQ -Upper Outer Quadrant

A -Adenine

G -Guanine

C -Cytosine

T -Thymine

CHAPTER 1

INTRODUCTION

Cancer is a diverse range of diseases that can start in any organ or tissue of the body when abnormal cells multiply uncontrollably and has the potential to surpass boundaries, spreading to different organs (metastasis). Cancer of the breast, lung and cervix are the top 3 cancers in Women in India in the year 2020. According to data collected by Population Based Cancer Research (PBCR) between 2010-2014, it is the third most common cancer among women in Mizoram. BC is the second most common cancer worldwide, comprising of about 10.4% of all cancers and is the second most common cause of cancer death in females (Abdullahi and Etemadi, 2016). Breast cancer (BC) is a malignant tumor that starts in the breast cells and is typically seen in the lobules or ducts of the breast. Rarely, connective tissues can also develop into a site of breast cancer. When cancer cells leave the lymph node and spread to nearby and distant healthy breast tissues as well as other parts of the body, this is referred to as metastasis. The distance that cancer cells have spread from the tumor's origin determines the stage of breast cancer. The hormonal receptor expression in breast cancer (BC) is lower and the age at presentation is younger among Indians as compared to western population (Sofi et al., 2012).

Breast cancer is a multifaceted and intricate illness that is influenced by a range of genetic, hormonal, and environmental variables, such as lifestyle choices, food habits, and reproductive history. Breast cancers are typically referred to as carcinomas, while they can also occasionally be called adenocarcinomas. Breast cancers can be classified into many categories based on the place and severity of the malignancy. "*In situ*" cancer means that it is still present at its original location; and "invasive or infiltrating" means that it has spread to neighboring tissues. The site of the cancer's onset was indicated by the names of the tissues or cells.

Breast cancer that is triple-negative or basal-like, lacks expression of the HER2/neu oncoprotein (HER2-) and the hormone receptor (ER-/PR-). BRCA1 gene mutations and younger women are commonly associated with triple negative breast

cancer. Breast cancer is a heterogeneous disease characterized by dysregulation of multiple cellular pathways and have different sensitivities to treatment (Liu et al., 2012). It is a heterogeneous disease with distinct biological subtypes and therefore, no gold standard therapy exists suitable for all tumours of the breast (Dawood, 2010; de Ruijter et al., 2011).

Immunohistochemistry (IHC) for receptors of estrogen (ER), progesterone (PR), and human epidermal growth factor 2 (HER2) expression is analyzed on all breast tumour specimens which will decide the type of treatment that the patients will be receiving. Triple negative breast cancer (TNBC) is an aggressive subtype having distinct clinical and biological characteristic defined by absence of ER, PR, and HER2 expression (Shimelis et al., 2018; Mousavi et al., 2019). Immunohistochemistry (IHC) is used to differentiate intracellular proteins or cell surfaces in tissues. IHC are utilized as prognostic and predictive markers as well as to inform therapy choices by dividing breast cancer into subtypes that exhibit distinct biological characteristics.

Approximately 10 - 15% of breast carcinomas are known to be of the TNBC subtype (Abdullahi and Etemadi, 2016), but may vary among different races, populations or geographic regions. Risk factors for TNBC include young age at breast cancer diagnosis, young age at menarche, young age at time of first child birth, high parity, lack or shorter duration of breast feeding, premenopausal women with high body mass index and African American ethnicity and an elevated waist:hip ratio (Dawood, 2010). TNBC is associated with advanced disease stage and higher-grade tumors at diagnosis and is associated with poor prognosis, as defined by low five-year survival and high recurrence rates after adjuvant therapy relative to other breast cancers subtypes (Shimelis et al., 2018; de Ruijter et al., 2011). Regardless of tumor size or lymph node status, triple negativity is an independent negative prognostic factor; reflecting its aggressive nature (Mousavi et al., 2019). TNBC has distinct clinical and pathologic features and is a clinical problem because of its relatively poor prognosis, aggressive behaviour and lack of targeted therapies, leaving

chemotherapy as the mainstay of treatment as it is resistant to current HER^{2-} targeted therapies and hormonal therapies (Irvin and Carey, 2008). At diagnosis, TNBCs are commonly of larger tumour size, inherently aggressive disease phenotype, have a higher histologic grade, high nuclear grade, high mitotic index, scant stromal content, central necrosis, pushing margins of invasion, a stromal lymphocytic response and multiple apoptotic cells (de Ruijter et al., 2011; Liu et al.,2012; Irvin and Carey, 2008). Histologically they are predominantly ductal, but other uncommon histologies including metaplastic, atypical or typical medullary or adenoid cystic carcinomas are also seen (Mousavi et al., 2019).

A small percentage of instances of breast cancer (5–10%) are inherited, and they are caused by mutations in the cancer susceptibility genes, BRCA1 and BRCA2, which are inherited autosomally. Just 10 - 20% of hereditary cases of breast cancer are caused by mutations in these two genes. There have also been reports of other genes, including TP53, PTEN, ATM, CHD1, CHEK2, SKT11, BRIP1, PALB2 and others having mutations that raise the risk of breast cancer. The hallmark of inherited breast cancer is often a younger age of cancer onset (~40 years) (Jian et al., 2017).

About 90% of incidences of breast cancer are sporadic. A number of reproductive, environmental, and demographic variables, in addition to those modifier genetic variants with a minor increase in risk or those predicted low-to-moderate penetrance, are important in the development of breast cancer. The cancer usually manifests as unilateral and has a late age of initiation (Sirisena et al., 2018). Breast cancer is a complex and heterogeneous disease resulting from uncontrolled cell division and growth. It is the leading cancer among women worldwide, with an annual incidence that has steadily increased over the years. Several factors contribute to the development of breast cancer including genetic, hormonal, environmental, and lifestyle factors. The development of breast cancer usually begins in the glandular cells that produce milk, or in the ducts that transport milk to the nipple. Mutations in these cells become damaged that disrupt their normal growth control mechanisms, resulting in the development of malignant tumours (Loeb and Loeb, 2000). Breast

cancer formation is a multi-step process involving various genetic and environmental factors.

Breast cancer can be classified into various subtypes based on HER2, PRs, and ERs. These subtypes exhibit varying treatment responses and prognosis. Among them, triple-negative breast cancer (TNBC) stands out as the most aggressive subtype, primarily because it lacks the expression of ER, PR, and HER2 proteins. Although TNBC represents approximately 10-20% of all breast cancer cases, it poses a disproportionately higher risk of mortality due to its aggressive behaviour and limited treatment alternatives. Consequently, gaining insights into the genetic characteristics of TNBC assumes great significance in the pursuit of targeted therapies aimed at enhancing patient outcomes (Landry et al., 2022).

Mizoram, a state in the north-eastern region of India with a unique socio-cultural context, provides an excellent environment for studying breast cancer epidemiology. Furthermore, Mizoram had the highest incidence of cancer in India between 2003 and 2010, but this dropped to fourth in 2012-2014, according to The Mizoram Population-Based Cancer Registry (PBCR) which reported cancer cases in 2003 at the regional level was established for event monitoring (Zomawia et al., 2023; Bhutia et al., 2020).

CHAPTER 2

REVIEW OF LITERATURE

Globally, cancer is one of the leading causes of death. Malignancy led to 8 million deaths in 2008; and by 2030, this number is projected to increase to 11 million. Because the majority of Asian nations are low- to middle-income nations, breast cancer is one of the major causes of mortality in these regions (Momenimovahed and Salehiniya, 2019).

Breast cancer (BC) is the common cancer in women and a main global health problem. India, a diverse and densely populated country, bears a high burden of breast cancer, with a large proportion of these cases being TNBC. It has become one of the most prevalent cancers among Indian women, with its incidence steadily increasing and prevalence rates peaking in recent decades. Several factors contribute to this increase in breast cancer cases among Indian women, including urbanization, changing lifestyles, delayed childbirths, and reduced breastfeeding, all of which increase the vulnerability of Indian women to breast cancer (Zodinpuii et al., 2022).

Mizoram has always been one of the hotspot regions for any type of cancer because of certain factors such as genotoxic stress from tobacco exposure, environmental or dietary risk factors that may contribute to or hasten the development (Ghatak et al., 2014; Yadav et al., 2018). It is very important to understand the contribution of different genetic factors for Breast Cancer in any given population. Even though Mizoram has second highest incidence of cancer in India, only few limited research in cancer related genes regarding mutational analysis have been done. This study has an objective of finding out the incidence of mutation rate and its role in the incidence of breast cancer among Mizo women.

According to India's National Health Policy 2017, DALY estimation is an essential epidemiological tool that should be used to evaluate epidemiological transitions, analyze macrolevel policies regarding health care use, assess the effectiveness of prevention and mitigation activities, allocate resources, and assess the nation's progress. With 1428 DALYs (DAYs - disability-adjusted life years) per

100,000 people, the northeast of India had the highest cancer burden (Kulothungan et al., 2022). The south and central areas of India had 1353 and 1351 DALYs per 100,000 people, respectively. Mizoram has the highest cancer-related DALYs per 100,000 people out of all the states/UTs, and in Delhi (2651 DALYsAMI per 100,000), Meghalaya (2609 DALYsAMI per 100,000). The greatest YLDsAMI from cancer was recorded in Mizoram (153 per 100,000) and Arunachal Pradesh (140 per 100,000) (AMI: adjusted mortality to incidence; YLD: years lived with disability). Within the nation, different regions have different cancer rates. According to the present survey, the north-east and northern regions had the highest cancer rates. The highest burden was found in the northern and northeastern regions, particularly with breast cancer in women and lung and esophageal cancer in men. The epidemiological transition levels within India's states are heterogeneous, and this is best captured by the nations within a nation description (Zomawia et al., 2023; Kulothungan et al., 2022).

Breast is a sexual feature of females and is the nutritional source for neonate and is the site for malignant change in one in ten women. They are also present in a rudimentary form in males. The incidence and mortality rates of breast cancer vary among races and ethnic groups worldwide, indicating that environmental and demographic factors, which are more significant and cannot be solely attributed to genetics, may influence the known risk factors for breast cancer (Lodha et al., 2011). Women in Saudi Arabia who marry younger have a significant correlation with breast cancer; those who marry before the age of 18 have a 13.9-fold increased chance of developing the disease. According to Alghamdi et al. (2015), the cause of the increased risk might be attributed to irregular or increased ovarian hormone output, such as estrogen in young girls. It has been discovered that the lifestyles of South Indian and Western women are associated with lower parity or fewer children, which increases the risk of breast cancer. Three or more live births are linked to a 40–50% risk decrease. Women who are alone or not married are more likely to have breast cancer, and it typically strikes in its advanced stages (Malvia et al., 2017).

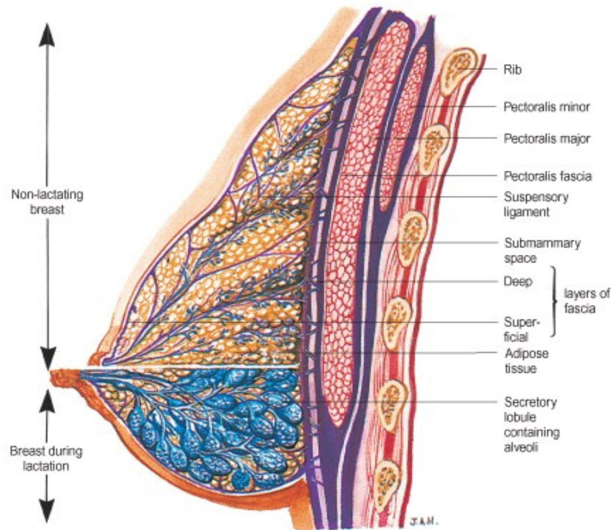


Figure 1. Macroscopic structure of the breast and changes during lactation
© Elsevier Ltd. 2005. Standring: Gray's Anatomy 39e-
www.graysanatomyonline.com

The deep pectoral fascia, which covers the pectoralis major, serratus anterior, inferior external oblique, and its aponeurosis, is where the breast rests (Figure 1).

The cancer hallmarks comprise of 6 biological functions acquired during development of tumours: proliferative signaling, evading growth suppressor, repelling cell death, aiding replicative immortality, stimulating angiogenesis, invasion and metastasis. Another two points was added in the last decade which includes reprogramming of energy metabolism and evading immune destruction.

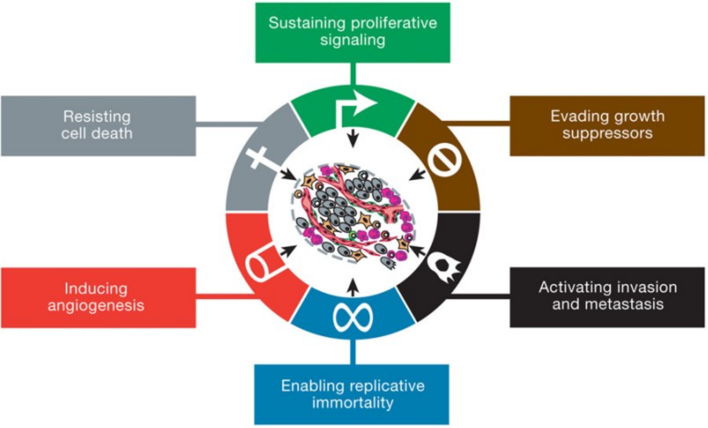


Figure 2. Hallmarks of Cancer (Hanahan et al., 2011)

©Elsevier Ltd. 2011

The body mass index (BMI), health, reproductive history (age at menarche, pregnancy, number of births, age at each birth, breastfeeding), socioeconomic status, smoking and alcohol habits, use of hormones (e.g., oral contraceptives), maximum attained weight and personal history of cancer are important parameters to assess the hereditary cancers (Díaz-Velásquez et al., 2023). Reproductive, hormonal, anthropometric, lifestyle, imaging, and many other factors are known to be related to the risk of developing breast cancer (BC), despite the fact that susceptibility to the disease is complex. It would be easier to identify people who are more likely to benefit and who are at higher risk, if comprehensive risk models consider all known characteristics (Lee et al., 2019).

According to Malvia et al. (2017), there is a correlation between chewing tobacco and betel quid chewing behaviors and breast cancer, which is the most common malignancy among women. While the chance of developing breast cancer is greatly enhanced, there is only a weak correlation between smoking and the illness. Women with a family history of breast cancer and those who began smoking in their early teenage or peri-menarchal years are at a heightened risk of developing breast cancer due to smoking (Jones et al., 2017). Many epidemiologic studies have shown that moderate alcohol use is associated with a 30 - 50% higher risk of breast cancer. A higher alcohol intake is associated with a higher risk of breast cancer. Because alcohol modifies estrogen levels, it affects breast density (Malvia et al., 2017).

It is commonly known that hormones and reproductive systems play a major role in the development of breast cancer. Due to cultural variations in breastfeeding, early age at first delivery, and parity- all of which are more common in Indian women than in Western women- the contributions of reproductive variables to the development of breast cancer in the two populations are extremely different. According to Palachandra et al. (2017), the main reproductive risk factors in industrialized nations include nulliparity, delayed age at first birth, and absence of breastfeeding.

Mammography density (MD), reproductive factors (e.g., age at menarche, age at menopause, parity, and age at first live birth), exogenous hormonal factors

(e.g., use of oral contraceptive [OC] and postmenopausal hormone replacement therapy [HRT]), anthropometric factors (e.g., height and body mass index [BMI]), and lifestyle factors (e.g., alcohol intake) are among the other risk factors that have been associated to an increased risk of breast cancer (Lee et al., 2019).

According to the 2018 Mizoram Population Census, Mizoram has the greatest prevalence of cancer, with the breast ranking between the top sites among women. Since breast cancer is one of the most common cancers, research on the genetic variations, frequency, and risk factors for breast cancer (BC) in this Mongolian population is still insufficient, and the findings may also apply to other populations (Zodinpuii et al., 2020). They studied the potential influence of various reproductive factors (10-30%), including parity, age at menarche, age at first delivery, number of live births, length of lactation, age at diagnosis, marital status, menopausal status, breastfeeding, use of oral contraceptive pills, abortion and tumor grade (Zodinpuii et al., 2020).

Family history is an established, significant causative factor for breast cancer. Women who have first-degree relative afflicted by breast cancer are twice as likely to develop the disease. The number of afflicted relatives and the age at diagnosis (less than 50 years) of first-degree relatives are associated with an increased familial relative risk (Brewer et al., 2017; Pachuau et al., 2022). The primary cause of a family history of breast cancer is inheritance of a high penetrance gene mutation in either BRCA1 or BRCA2, but these genes only account for 10 to 20% of cases of familial breast cancer, indicating the possibility that other high, moderate, and low penetrance genes also contribute to breast cancer susceptibility (Jian et al., 2017).

Evaluation of the histological tumor grade for the tubular differentiation composite, nuclear features, and mitotic activity is one of the most crucial factors in determining the type of breast cancer. Tumor grade is a significant prognostic factor, especially in cases of early-stage breast cancer, where axillary lymph nodes are not involved or are involved seldom (Suba, 2014). Hormone receptors and HER2neu status have been found to positively correlate with tumor grade; moreover, an inverse association between ER/PR/HER2neu and tumor grade was established, irrespective

of geographical variations and ethnicity. It was discovered that when tumor grade increases, so does the expression of these hormone receptors and HER2neu oncoprotein (Yin et al., 2018).

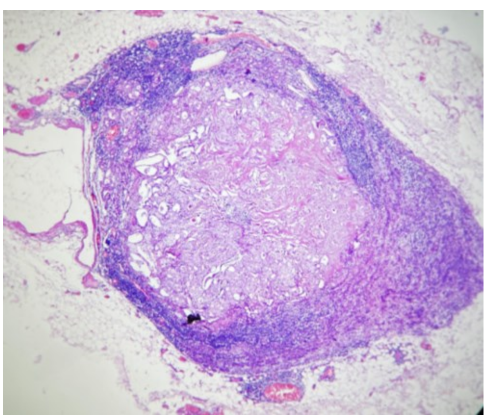


Figure 3. Tumor metastasis in Lymph node.

Axillary lymph nodes receive 75% of the lymph of the breast and forms a usual component of the breast surgery and is executed because of metastases and axillary lymph nodes have great prognostic significance and guides adjuvant therapy (Figure 3).

TNBC is defined by lack of expression of the ER, PR and HER2neu. It is a subgroup with a distinct clinical and biological characteristic. TNBC consists about 10 to 20 % of all breast cancers worldwide (Mousavi et al., 2019; Yao-Lung, 2011) and it is estimated that 25% of TNBCs carry *BRCA1* mutation. More than 75% of tumors with *BRCA* mutation are triple-negative. There is also high genomic instability and mutations or loss of genes *TP53, PIK3CA, PTEN, RB1, BRCA1*, low *Bcl-2* and high *Ki67* expression (Liu et al., 2012; Chivukula et al., 2008). TNBC responds well to chemotherapy, but its prognosis, OS (overall survival), and DFS (disease free survival) are worse due to the absence of target treatment in the follow-up (Mousavi et al. 2019). *HER2* (c-erbB-2 and neu) is a member of type 1 TK family and is an oncoprotein overexpressed (20%) in invasive primary breast cancer. HER2/neu overexpression is an independent prognostic and predictor factor in breast carcinoma (Moinfar, 2007).

Studies from the Indian Genome Variation Consortium (IGVC, 2008) has suggested that the genetic basis of diseases in the Indian subcontinent to be distinct. Therefore, it is essential to create a somatic mutation landscape of breast cancer at whole exome level. Genetic profiling of each population is expected to play a major role in disease diagnosis, adverse drug response and drug treatment. The cellular and molecular heterogeneity of breast tumors and the large number of genes potentially involved in controlling cell growth, death and differentiation emphasize the importance of studying multiple genetic alterations (Sørlie et al., 2001). Mutation frequency of genes differ greatly among populations from different geographic regions and ethnicities (Somasundaram, 2010). About 25% of TNBCs have *BRCA1* mutation and 75% of tumors with *BRCA* mutations are triple-negative. Some of the risk factors for TNBC are BRCA mutation, ethnicity, age and Body Mass Index (BMI) (Mousavi et al., 2019). TNBC has a worse prognosis, high recurrence and poor survival rates when compared to other subtypes of breast cancer (Dawood, 2010). There is also high genomic instability and mutations or loss of genes TP53, PIK3CA, PTEN, RB1, and BRCA1, low Bcl-2 high Ki67 expression (Liu et al.,2012; Chivukula et al., 2008). With the reducing cost of DNA sequencing and increased efficiency, whole exome sequencing has replaced a considerable number of genetic screening assays. Till date there are only few studies in which whole exome or whole genome sequencing has been performed.

Rare, high penetrance variations in BRCA1 and BRCA2; five rare, intermediate-risk variants (e.g., in PALB2, ATM, and CHEK2); and commoner variants (mainly single nucleotide polymorphisms [SNPs]) conferring lower risks constitute the genetic susceptibilities to BC (Lee et al., 2019).

The lower survival rate associated with TNBC is attributable to high genomic instability, high mitotic index, mutations in *p53* and *BRCA1*, lymphatic diffusion, not detectable with standard imaging and are detected in significant tumor size, lack of response to current breast cancer treatments for the clinical management of breast cancer i.e, hormonal and HER^2neu inhibitor therapy. Targeted therapeutic options are still under investigation for TNBC (Pierobon and Frankenfield, 2013; Whitesell et

al., 2014). The normal treatment protocol is surgery, chemotherapy, radiotherapy followed by adjuvant therapy (Somasundaram,2010). ER / PR positive tumors are treated by Tamoxifen. HER-2/neu positive tumors are treated by a monoclonal antibody. No adjuvant therapy is available for ER/PR and HER-2/neu negative (triple negative) tumors (Sinha et al., 2016).

The primary characteristic of breast cancer is a proliferation of uncontrolled cells due to genetic mutations that accumulate over time within the breast tissue. The key genetic events or events contributing to the development of breast cancer include:

(i) BRCA1 and BRCA2 Mutations: Mutations in the BRCA1 and BRCA2 genes pose a notable elevation in the risk of developing breast cancer. The genes are important for maintaining genomic stability and repairing DNA damage. Mutations in these genes can lead to faulty DNA repair mechanisms, which can lead to the accumulation of mutations and cancer (Miki et al., 1994). The penetrance of these mutations is estimated to be between 33% and 70%, influenced by variables like age at diagnosis and familial background. PALB2 has been identified as a common predisposing gene for breast cancer, with its penetrance varying based on individual factors (Siraj et al., 2023)

(ii) HER2 (Human Epidermal Growth Factor Receptor 2) Amplification: It is estimated that 20-25% of breast cancers overexpress the cell surface receptor HER2, which regulates cell proliferation and growth. Amplification of HER2 can cause aggressive breast cancers and uncontrolled cell proliferation. Furthermore, approximately 10-20% of invasive breast cancers (BC) are triple negative, which is defined by the lack of HER2 gene amplification, progesterone receptor (PR), and estrogen receptor (ER) immunohistochemistry (Angius et al., 2023).

(iii) TP53 gene Mutation: Tumour-suppressor gene that controls cell division and prevents tumour formation. *TP53* encodes the protein (43 KDa) that is most commonly found in the cell cytoplasm. It binds specifically to DNA, and post-translational changes such as phosphorylation, acetylation, methylation, and ubiquitination control its function. TP53 is an effective tumor suppressor gene that

has a wide range of uses. Mutations in the TP53 gene can disrupt this control, leading to cancer development. More than half of cancers contain a mutation in p53, which inhibits both immune escape and innate immunity (Sha et al., 2022).

(iv) Hormone receptor status: Estrogen and progesterone receptors play a key role in cell growth in breast cancer cells. The hormone therapy is effective against tumours that are hormone receptor-positive and are dependent upon these hormones to grow. It has been shown that selective estrogen receptor modulators like raloxifene and tamoxifen, as well as aromatase inhibitors like anastrozole and exemestane decrease the incidence of breast cancer by 30% to 60% in randomized placebo-controlled trials for women with elevated breast cancer risk (genetic status was only available for very few of these women). This reduction is especially significant in cases that are estrogen receptor-positive (Hu et al., 2021).

Whole Exome Sequencing (WES) has emerged as a powerful tool for comprehensively profiling the genetic alterations in TNBC. By sequencing the exons, which are the protein-coding regions of the genome, WES allows for the identification of both DNA variations at somatic mutations and germline level. This high-resolution approach has been instrumental in uncovering the genetic drivers and mutational signatures of TNBC, thereby paving the way for personalized medicine strategies. Through whole exome sequencing data analysis, several other genes associated with breast cancer that are frequently altered have been identified in previous studies. A study identified seven high-risk genes associated with hereditary cancer susceptibility in Ashkenazi Jewish (AJ) women in Mexico, including *APC, CHEK2, MSH2, BMPR1A, MEN1, MLH1,* and *MSH6*. 14% of those participants possessed pathogenic or likely pathogenic variants (Díaz -Velásquez et al., 2023). Moderate-high penetrance genes for Hereditary Breast Cancer in Romania, including *TP53, PALB2, PTEN, CDH1, ATM* and *CHEK2* was identified (Catana et al., 2023). *PIK3CA, TP53, GATA3, MAP3K1, CDH1, CBFB, PTEN*, and *RUNX1* significantly mutated genes, and confirmed somatic mutations in *AKT1, BARD1, MAP3K1*, and *MET* using sanger sequencing (Ding et al., 2023). Dorling et al. (2021) found that over 60,000 breast cancer patients and over 53,000 healthy women took part in

research conducted by the (BCAC) and were able to sequence their germline DNA. In their study, it was discovered that protein-truncating mutations in five specific genes, namely *ATM, BRCA1, BRCA2, CHEK2,* and *PALB2*, were linked to substantial risks of developing breast cancer. Seven other genes showed weaker evidence of a relationship with breast cancer, including *BARD1, RAD51C, RAD51D, PTEN, NF1, TP53*, and *MSH6.*

Mizoram, the north-eastern state of India, exhibits a unique demographic and genetic profile as compared to the rest of the country. The majority of Mizoram's population consists of indigenous groups with distinct genetic, lifestyle and environmental backgrounds. These differences may reflect differences in the genetic and epidemiological patterns of TNBC in Mizoram as compared to the rest of the other Indian population. Moreover, the genetic status of TNBC in the Indian population, including specific subpopulations such as Mizoram population, has still not been thoroughly investigated. As India has a large genetic diversity and ethnicity may influence TNBC genetics, identifying genomic variation in TNBC in the Indian context specifically in Mizoram will be vital in developing targeted therapies and their pathogenicity for TNBC. Therefore, studying TNBC specifically in Mizoram offers valuable insights into regional and ethnic disparities in this disease.

This study aims to address the genetic landscape of TNBC in India, with a particular emphasis on the state of Mizoram. A WES analysis of somatic tissue TNBC samples was first conducted, followed by a germline variant analysis. Our study encompassed a comprehensive genomics analysis, including Single Nucleotide Variants (SNV) including Single Nucleotide Polymorphism (SNPs) and Indels, identification of frequently mutated genes, exploration of mutually exclusive genes, analysis of mutational signatures, assessment of drug-gene interactions, and an in-depth investigation of pathways affected by these genetic variations. This comprehensive analysis was followed by an in-depth investigation into germline variants. We also conducted a clinical data analysis of somatic TNBC samples to gain crucial insights into the disease. Through this process, we examined the clinical data associated with somatic TNBC samples and identified variants related to their

clinical characteristics. An analysis of TNBC from both a genetic and a clinical perspective sheds light on how genetic mutations and clinical manifestations impact TNBC and identify genes and variants associated with TNBC. Therefore, this study provides a deeper understanding of the genetic factors underlying TNBC in Indian populations, particularly in Mizoram. Consequently, this study may provide a crucial contribution towards identifying the possible causes of the disease and may aid in developing target therapies and clinical improvements for the treatment of TNBC.

The study aims to analyze the Clinicopathological and hormonal status of the TNBC cases. Further, the genomic changes were screened through whole exome analysis. The findings of this study may also provide useful information on the prevalence of mutations and non-genetic factors as risk factors for the development of breast cancer in Mizo population.

CHAPTER 3

AIMS AND OBJECTIVES

The objectives of the study were:

1) To analyze Triple Negative Breast Cancer (TNBC) with clinico-pathological parameters.
2) To evaluate the ER, PR and HER-2/neu expression in invasive breast carcinomas.
3) To perform whole exome analysis to characterize the landscape of genetic alterations underlying TNBC in Mizo population.

CHAPTER 4

MATERIALS AND METHODS

Patients who had undergone modified radical mastectomy at Civil Hospital Aizawl were included in the study. The histological type, involvement of lymph node, tumor grade, ER and PR receptors and HER-2/neu status were investigated for all the samples. Fresh as well as formalin fixed paraffin embedded (FFPE) breast tissue from the main tumour and adjacent normal tissue and corresponding blood samples was taken for analysis. Diagnosed tumour tissue where the only sample available is core biopsy, incisional, excisional, lumpectomy or cytology were excluded from the study as adjacent normal tissue sample and lymph nodes were not be available. A skilled technician extracted 2 ml of blood from the participants (patients and controls), the blood was placed in EDTA vials and stored at -20°C for further analysis.

Clinical records, age at diagnosis, gender, tumor type or grade, habits and familial history of the patients were collected through structured questionnaire. Samples were collected only with prior consent of the patients. A total of 240 samples of MRM specimen were collected and 59 patients are found to be Triple negative in immunohistochemistry and were included in the study. Whole Exome Sequencing (WES) was done on 15 patients tumour, adjacent normal and whole blood sample.

The demography, reproductive history, environmental variables, and family history of the recruited individuals with regard to cancer and other hereditary disorders are among the epidemiological aspects taken into consideration. The patient's reproductive history was taken into consideration, including their age at marriage/ first delivery / menarche/ menopause, parity, marital status, no. of children, length of breastfeeding, usage of oral contraceptive pills, and history of abortion. Environmental factors include eating habits (fruits, vegetables, *sa-um*, smoked food, salt, water, and oil), sleeping schedules, night shift work, and exercise habits. Tobacco and alcohol intake are also factors (Betel nut, gutkha, sahdah, khaini, tuibur and cigarette).

Subject inclusion criteria

1. All patients with breast cancer undergoing Modified radical mastectomy regardless of presence or absence of similar tumour in the family.
2. Not associated with any kind of chronic diseases.
3. Clinically diagnosed by Surgeon and confirmed by Pathologist.
4. Willingness to participate in the study.

Subject exclusion criteria

1. Breast cancer patients without Modified radical mastectomy.
2. Patient having any other Chronic/ major illness
3. Patients who were pre-treated for any other type of cancer.
4. Unwilling patients.

Age-based frequency distribution of the study's cases and controls was computed. The relationship between demographic characteristics and TNBC risk was evaluated using chi-square testing. To calculate the possible confounder's impact of environmental variables on breast cancer, logistic regression analysis was performed. IBM Statistical Package for Social Sciences (SPSS), v22 was used. The odds ratio and confidence interval were calculated using MedCalc v20.113 (https://www.medcalc.org/calc/odds_ratio.php).

Ethical Clearance

Ethical approval for the study was obtained from Civil Hospital Aizawl (B.12018/1/13-CH(A)/IEC/33) and Mizoram State Cancer Institute- Human Ethical Committee.

Clinico-pathological parameters like patient's age, histological type, grade and lymph node metastasis

On microscopic examination, the histological type and tumour grade as per Modified Bloom Richardson score would be classified. In this scoring system, three

factors that are taken into consideration and each of these factors are scored from 1-3 (Sinha et al., 2016).

- **The amount of gland formation or differentiation of the tumor: A**ssessed over the whole tumour and is a low/medium power assessment.

Score 1: >75% of tumor area forming glandular/tubular structures
Score 2: 10% to 75% of tumor area forming glandular/tubular structures
Score 3: <10% of tumor area forming glandular/tubular structures

- **The nuclear features or nuclear pleomorphism**: Done on the worst area

Score 1: Tumour cells are slightly larger in size in comparison with normal epithelial cells, have small nuclei, uniform nuclear chromatin, regular outlines and little variation in size
Score 2: Tumour cells are larger than normal, have open vesicular nuclei, visible nucleoli and moderate pleomorphism in both size and shape
Score 3: Tumour cells have vesicular nuclei, often with prominent nucleoli, exhibiting marked pleomorphism in size and shape, occasionally with very large and bizarre forms.

- **Mitotic activity of tumor** (considering field diameter of microscope to be 0.58 mm and area to be 0.264 sq.mm.): Done in the highest mitotic area.

Score 1: less than or equal to 9 mitoses/ 10 high power fields
Score 2: 10-19 mitoses/ 10 high power fields
Score 3: equal to or greater than 20 mitoses/ 10 high power fields.
Each score will be addedup to give a final total score (3-9). The final total score will be used to determine the grade in the following way:

Grade 1 tumors have a score of 3-5
Grade 2 tumors have a score of 6-7
Grade 3 tumors have a score of 8-9

Presence of lymphovascular invasion, perineurial invasion and in situ component would also be determined (Sinha et al., 2016).

Hormonal receptor expression by IHC

Modified radical mastectomy specimens from Surgery operation theatre, Civil Hospital Aizawl were grossed as per standard protocols and histopathological examination by haematoxylin-eosin staining for microscopic examination was done. Tumor characteristics regarding type of tumor and histological grade are classified as per modified Bloom Richardson histologic score which take into account the amount of gland formation, nuclear features or nuclear pleomorphism and mitotic activity of the tumor. ER, PR scoring will be done as per the Allred scoring system and HER-2/neu scoring will be done according to the standard reporting protocols.

Immunohisto Chemistry Protocol

The formalin fixed paraffin embedded block having tumour with adjacent normal breast tissue, which will serve as internal control, were selected to perform IHC. Appropriate external control was also selected. About 3.5-4.0 µm sections were made and spread into poly-L-lysine coated slides and De-paraffinized in an oven at 60ºC overnight**.** It was transferred to xylene for 20 minutes for clearing. Rehydration were achieved by immersing in decreasing concentrations of alcohol- 100%, 70% and 50% for 5 mins each and washed in running water for 5 mins. For Antigen retrieval, slides were kept at room temperature for 5 mins, and then immersed in a solution bake at 60ºC for 10 minutes [6.05g Tris (hydroxyl methyl) aminomethane salt (Merck) and 0.744g EDTA dipotassium salt dehydrate GR (Merck) powder dissolved in 1 litre distilled water with pH 9] and washed in buffer [6g Tris (hydroxyl methyl) aminomethane salt (Merck) dissolved in 1 litre distilled water with pH 7.4]. After that, the Primary antibody (Cell marque) was added to the slide and incubated in humidifier for 10 mins and then washed in wash buffer for 3 mins. Next, Superblock were added to slide and incubated in humidifier for 10 mins and washed with buffer for 3 mins. HRP polymer were added to slide and incubated in wet chamber for 30 mins. After incubation, the slide was washed with buffer for 3 mins. For staining, Dab was used and incubated in wet chamber until desired color reaction

is observed. Counterstained with Hematoxylin for 1 minute and washed in running water for 5 mins.. The slide was dehydrated in increasing concentration of alcohol 50%, 70% and 100% for 1 min each and air dried. The slide was mounted in DPX (Dibutylphthalate polystyrene xylene). Result were interpreted in the tumour area and invasive tumour area for Her2neu. Reports are given only when internal control is positive for hormone receptor and external control is positive for Her2neu. Immunoreactivity as per the ASCO-CAP guidelines (2018) of ER/ PR and HER2 was done.

Allred scoring

Proportion score:

0 - No cells are ER/PR positive

1 - ≤ 1% of cells are ER/PR positive

2 - 1-10% of cells are ER/PR positive

3 - 11-33% of cells are ER/PR positive

4 - 34-66% of cells are ER/PR positive

5 - 67-100% of cells are ER/PR positive

Intensity score:

0 – Negative

1 – Weak

2 – Intermediate

3 – Strong

HER-2/neu scoring was done according to the standard reporting protocols

Score to Report	**HER2/neu protein Assessment**	**Staining pattern**
0	Negative	No staining is seen or membrane staining is seen in <10% of invasive tumor cells
1+	Negative	Faint/Barely perceptible membrane staining detected in >10% of invasive tumor cells

2+	Equivocal	Weak to moderate complete membrane staining in >10% of invasive tumor cells
3+	Positive	Strong complete membrane staining in >30% of invasive tumor cells.

Whole Exome Sequencing and data analysis

DNA isolation Using QIAamp AllPrep DNA Mini Kit from Tissue Samples

Genomic DNA was isolated from frozen tumor tissues using AllPrep DNA Mini Kit (QIAGEN, Lot. 51304). 30 mg of tissue is placed in 1.5 ml microcentrifuge tube and 600 µl of Buffer RLT Plus added, homogenized the lysate using pestle until the tissue lyase. The lysate underwent centrifugation for 3 minutes at 14,000 rpm, and the supernatant was carefully aspirated using a pipette. The homogenized lysate was then transferred to an AllPrep DNA spin column, positioned within a 2 ml collection tube. After centrifuging for 30 seconds at 10,000 rpm, 500 µl of AW1 Buffer was introduced to the AllPrep DNA spin column. A subsequent 15-second centrifugation at 10,000 rpm facilitated the washing of the spin column membrane. The flow-through was discarded, and the AllPrep DNA spin column was relocated to a fresh 2 ml collection tube. To further cleanse the spin column membrane, 500 µl of AW2 Buffer was added, and the lid was closed gently. This was followed by a 2-minute centrifugation at 14,000 rpm. Finally, the AllPrep DNA spin column was transferred to a new 1.5 ml Eppendorf Microcentrifuge Tube, and 70 µl of BE buffer was directly added to the spin column membrane, allowing it to incubate at room temperature for 5 minutes. Centrifuged for 1 minute at 10,000 rpm to elute the DNA, stored at -20°C until use for Whole Exome Sequencing.

DNA isolation Using QIAamp® Blood Mini Kit from Blood Samples

Genomic DNA from Whole Blood was isolated using QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN) and stored in -20°C. 20 µl of QIAGEN Protease was

pipetted into 1.5 ml microcentrifuge tube. Mix 200 µl of the sample thoroughly with 200 µl of Buffer AL by vortexing. Incubate the mixture at 56°C for 10 minutes. After the incubation, release and invert the tubes five times. Briefly centrifuge the 1.5ml microcentrifuge tube to eliminate drops from the lid. Add 200 µl of CHILLED ethanol, mix thoroughly by vortexing, and again briefly centrifuge the 1.5 ml microcentrifuge tube to remove drops from the lid. Transfer the entire lysate into a QIAamp mini-Spin column placed in a 2 ml collection tube (both column and tube should be labeled) using a pipette. Centrifuge the column at room temperature at 8000 rpm for 1 minute. Discard the flow-through and the collection tube, then place the spin column in a new 2 ml collection tube. Add 500 µl of Buffer AW1 into the QIAamp mini spin column, and centrifuge at 8000 rpm for 1 minute at room temperature. Discard the flow-through and the collection tube, and place the spin column back in a 2 ml new collection tube. Add 500 µl of Buffer AW2 into the spin column, and centrifuge at 14,000 rpm for 3 minutes at room temperature. Discard the flow-through and the collection tube, then place the column back in the same collection tube. Centrifugation was done at 14,000 rpm for 1 minute at room temperature (Blank Spin). Discarded the 2 ml collection tube and placed the column back in a new 1.5 ml microcentrifuge tube. The caps were opened and dried in the laminar air flow (LAF) / air dry for 15 minutes. Added 75 µl of AE buffer directly into the membrane of the column and incubated for 5 minutes at room temperature in LAF. Centrifuged for 8000 rpm for 1 minute at room temperature to elute the DNA. The elution was stored at -20°C and used for Whole Exome Sequencing.

Preparation for 0.8% agarose gel in TAE (Tris acetate- EDTA) Buffer (1X)

In 100 ml of Conical flask, 0.32 g agarose powder was added and 40 ml of TAE (Tris acetate- EDTA) buffer was used to dissolve, oven-heated and cooled down. 4 µl EtBr (Ethidium Bromide) was added to the luke warm gel and poured on the tray to solidify the gel. The agarose gel was placed in the electrophoresis chamber and 3 µl of Genomic DNA samples and 2 µl of 100bp ladder were loaded to the well and run for 30 minutes to check the quality and concentration for Genomic DNA.

Library Preparation for Exome Sequencing and Sequencing Run

High quality Genomic DNA was used for Exome sequencing library preparation using Illumina Truseq Exome Enrichment Kit (Illumina). Initially, 100 ng of genomic DNA underwent fragmentation using the Covaris (ME220) Instrument, resulting in blunt end fragments. Subsequently, the DNA fragments underwent end-repair and adenylation, followed by ligation with dual indexing (unique) primer pairs incorporating barcoded adapters. This was followed by limited cycle PCR amplification. The pooled amplified genomic DNA (gDNA) libraries were then subjected to hybridization with exome-based enrichment probes. Exonic regions were captured using streptavidin-coated magnetic beads, and a further round of limited cycle PCR amplification was conducted to generate enriched exome library pools. To assess the quality of the enriched exome libraries, analysis was performed using the Agilent 2200 TapeStation with high sensitivity D1000 ScreenTape. Finally, the pooled libraries were loaded onto the S4 Flow cell for a 2 X 100 bp paired-end sequencing run in the Illumina Novaseq 6000 sequencer.

Data quality check and alignment

Sequencing of 15 tissue cancer samples and adjacent normal samples from triple-negative breast cancer (TNBC) patients and 15 blood (germline) samples were done. The raw sequencing data were processed using Fastp in paired-end mode, with a PHRED score cut-off of 30 used to remove low-quality reads and adaptors. FastQC was used to assess the quality of raw reads and trimmed reads that exceeded 30 base pairs and had an error rate below 10% were retained. Trimmed reads were aligned to the human reference genome GRCh38 using default parameters in BWA. Aligned reads were sorted and indexed with Samtools, and duplicate marking was performed using Picard. Base quality score recalibration was done using Genome Analysis Toolkit (GATK) with known sites vcf files of dbsnp version 146 and Mills and 1000G gold standard indels with default settings and Post-quality control.

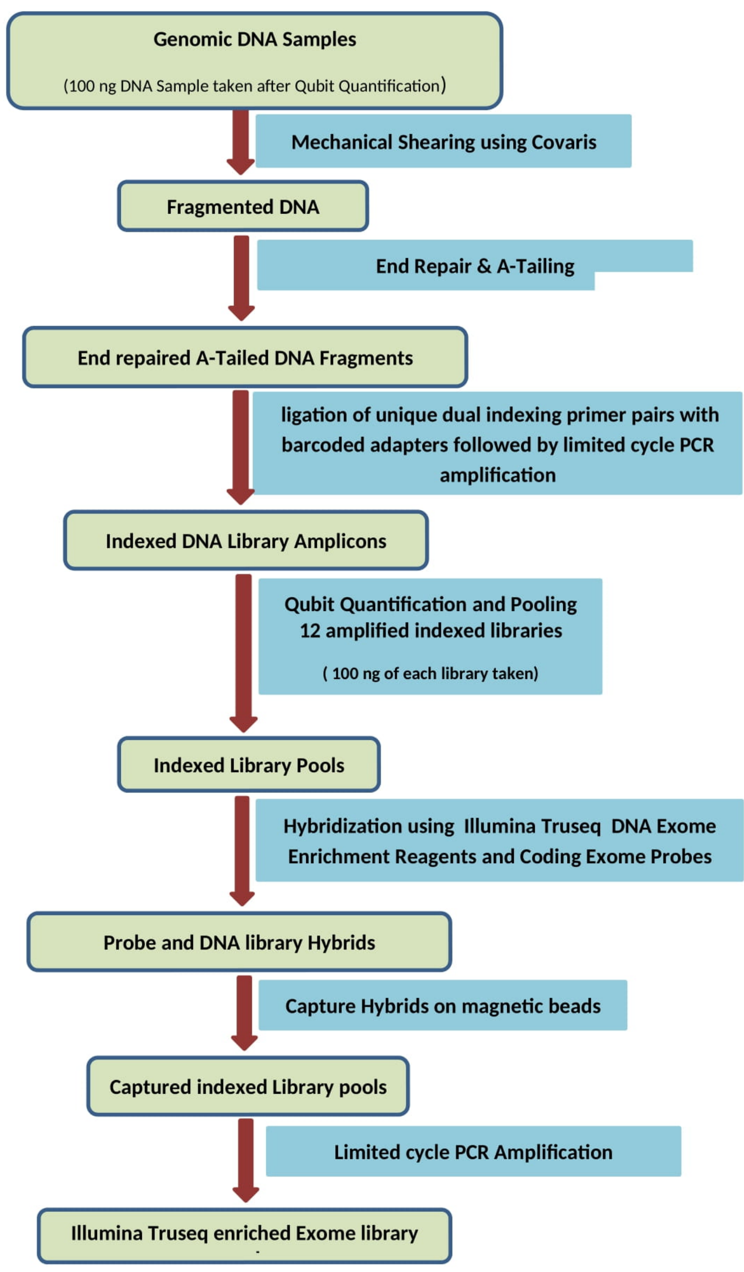


Figure 4. Workflow for library preparation for whole exome sequencing

Somatic variant calling

Somatic variants were called using Mutect2 with matched tumor/normal pair mode, a variant caller within the GATK workflow. The output bam file was processed with gatk GetPileupSummaries with known variants, integrating population allele frequencies of common and rare alleles from gnomAD, alongside a

bed file delineating exome intervals, effectively circumventing the calling of germline variants. Following this, the Mutect2-filtered variants were subjected to annotation via Ensembl Variant Effect Predictor (VEP). Further filtration based on contamination estimates was performed using GATK FilterMutectCalls, ensuring that only somatic variants meeting the filter criteria proceeded for subsequent analysis. The FILTER field is labelled with PASS for calls that are likely true positives, and 14 filters, including contamination, appear to be applied at this step.

Germline variant calling

The germline variants were called using GATK Haplotypecaller and annotated using Annovar. Following this, a manual filtering process, specifically focusing on exonic variants, non-synonymous and pathogenic variants, was carried out to select only the important variants. In order to further assess the functional impact of these variants, three different predictive tools, SIFT, PolyPhen2, and MutationTaster were utilized. In the subsequent analysis phase, only exonic nonsynonymous variations deemed deleterious by at least two of the following tools: SIFT (labeling as Damaging), PolyPhen2 (labeling as Probably Damaging), and MutationTaster2 (labeling as Disease causing) were selected for downstream analysis.

Somatic interactions

The identification of mutually exclusive or co-occurring pairs of mutated genes within our samples was conducted using the somaticInteractions feature integrated within Maftools, which utilizes Fisher's exact test (among the top 30 most mutated genes). In cancer, numerous disease-causing genes exhibit either co-occurring patterns or distinct exclusivity in their mutation behaviors. To identify such sets of genes, the somaticInteractions function in Maftools is employed, employing a pairwise Fisher's exact test to pinpoint significantly associated gene pairs. Furthermore, the somaticInteractions function includes the comet Exact Test to identify potentially altered gene sets that involve more than two genes. For this analysis, the top 50 driver genes were utilized, labelling the results with a threshold P value of 0.05 and 0.01. This process aids in unveiling exclusive or co-occurring

relationships among mutated genes, contributing to a better understanding of their roles in cancer development and progression (Source: OncoVar: an integrated database and analysis platform for oncogenic driver variants in cancers).

Oncogenic signaling pathway enrichment analysis

The Database for Annotation, Visualization and Integrated Discovery (DAVID) was used. This functions by computing the proportions of genes affected within a pathway, as well as determining the fractions of samples displaying pathogenetic or unknown variants in these pathway genes.

Mutational Signature analysis

Analysis of mutational signatures was conducted by utilizing our dataset sourced from filtered somatic variants in VCF files using Maftools in R v.4.3. The analysis method involved extracting neighboring bases adjacent to the mutated ones and creating a matrix for pathogenetic or unidentified variants. This matrix classified nucleotide substitutions into 96 classes based on the nearby bases, achieved using the *trinucleotideMatrix* function. Non-negative matrix factorization (NMF) was then employed, utilizing the *estimateSignatures* function to estimate the number of signatures. Three optimal signatures were determined based on the cophenetic correlation metric. The pathogenetic or unknown variant matrix was further divided into these three signatures with the *extractSignatures* function. These extracted signatures were compared to both previously known COSMIC signatures—30 legacy signatures and a more recent set of 65 Single Base Substitution (SBS) signatures—using the *compareSignatures* function. The resulting signatures were then visually represented using the *plotSignatures* function.

CHAPTER 5

RESULTS

The format used for Modified radical mastectomy in Civil Hospital Aizawl, reporting the type and grade of tumour, in situ component, nerve and lymph node involvement is given below:

HOSPITAL & MEDICAL EDUCATION
CIVIL HOSPITAL AIZAWL
xxxxxxxxxxxxx
PIN:796001
DEPARTMENT OF PATHOLOGY

HISTOPATHOLOGY
Phone No: 2316570, FAX:0389 2316570, 301 (HISTOPATHO)
Page 1 of 1

M.R NO	:xxxxxxx	**REPORT NO**	: xxxxxxxxxxx
INDOOR ADM. NO	:xxxxxxx	Date of Operation	: xxxxxxxxxxx
PATIENT'S NAME	:xxxxxxx	Sample Received On	:09/08/2023 00:00:00
SEX	:Female	Grossing Date	: xxxxxxxxxxx
AGE	:48 years	REPORT DATE	:16/08/2023 12:55
ADDRESS	:xxxxxx	**HISTOPATH NO.**	:xxxxxxx
Ref. by	:X	Report Printed on	:15/11/2023 09:39:31
Ref. Unit	:FEMALE SURGICAL WARD WARD	BED NO.:	:xxxxxx

HISTOPATHOLOGY REPORT (Test Date/Time: xxxxxxxxxxxxxxx)

TESTS	RESULT/OBSERVATION
Specimen :	Left breast, modified radical mastectomy
Gross Finding :	Specimen type: Left breast, Modified Radical Mastectomy Received a Left breast, Modified Radical Mastectomyspecimen, measuring 23 x14 x 6cm, with overlying skin flap measuring 15 x 5cm. The nipple and areola unremarkable. No linear scar is seen. On serial sectioning firm grey brown tumor is identified, measuring 3 x 2 x 2cm, located in the upper inner. It is located 1.5cm from the skin 1.5cm from the base. The remaining breast tissue shows fibrosis. Altogether 10 lymph nodes dissected, largest measuring 1.5cm diameter firm grey brown.
Report:	Specimen:Left breast, Modified Radical Mastectomy Invasive ductal carcinoma, Grade II Modified Bloom Richardson score : 2+2+2 = 6 Extensive intraductal component (EIC) is absent. No involvement of nipple or areola by either in situ or invasive carcinoma is identified. The base is free of tumour Lymphovascular emboli are seen Perineurial invasion are absent The overlying skin is unremarkable The adjacent breast show desmoplastic reaction. Lymph nodes : i)Attached axillary tail : 0/10 - Free of tumor. IMPRESSION : Specimen:Left breast, Modified Radical Mastectomy Invasive ductal carcinoma, Grade II, without nodal involvement.

Grossed by Reported by DR. C LALCHHANDAMA (MD, PATH.) DR. C LALCHHANDAMA (MD, PATH.) PATHOLOGIST PATHOLOGIST

Figure 5.a.b. shows 3+ nuclear positivity for ER and PR in 80% of tumour cells. Figure 5.c. shows 3+ complete membranous positivity in >10% of tumour cells. Figure 5.d.e.f. are all negative for ER, PR and Her2neu.

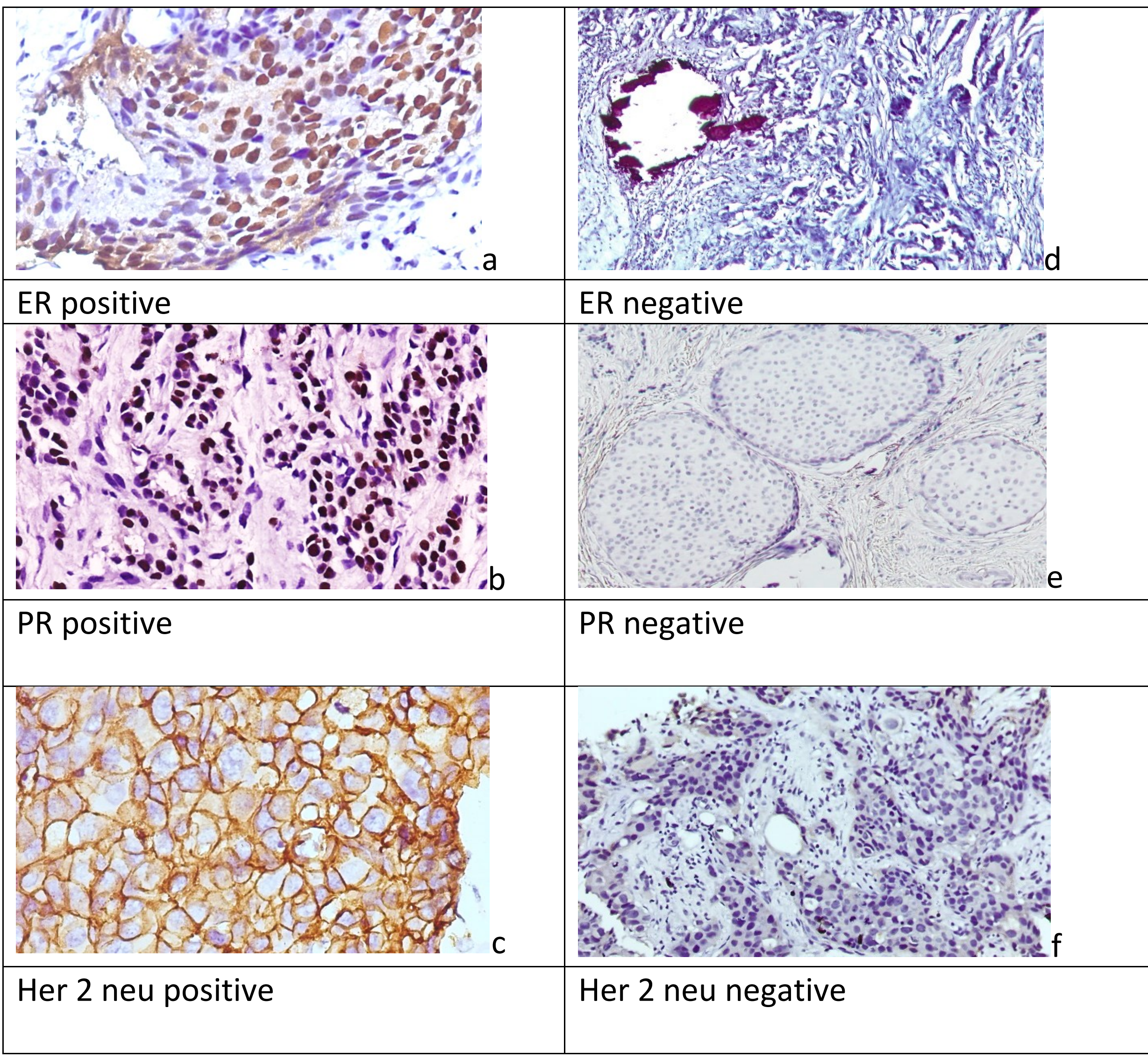


Figure 5. IHC slide images of representative samples

The study consisted of 59 cases, among them 39 (66.1%) cases > 45 years and 20 (33.9%) cases < 45 years. Chi-square goodness of fit test was used to compare the proportion of cases equal between the age group of cancer patients, frequency of exercise and sleep hrs. The results reveal that more cases were observed at the age group above 45 years (X2=6.119, P<0.05). Exercise, 40 (67.8%) of the cases did not do exercise and only 10 (16.9%) of cases did exercise every day (X2=31.559, P<0.05). The results show that 43 (72.9%) of the cases were having sleeping > 6 hrs and 27.1 % of the cases slept less than 6 hrs (X2=45.186, P<0.05) (Table 1).

Table 1. Frequency analysis of sociodemographic variables among the TNBC patients

Demographic Variables	Categories	Mean	SEM	Frequency (%)	Chi square Value	P Value
Age in Years		52.17	1.84			
BMI		23.71	0.49			
Age	<45 Years			20 (33.9)	6.119	**0.013**
	>45Years			39 (66.1)		
Exercise	Never			40 (67.8)	31.559	**0.000**
	Once a week			9 (15.3)		
	Everyday			10 (16.9)		
Sleep (hrs)	1-3 hrs			2 (3.4)	45.186	**0.000**
	3-6 hrs			14 (23.7)		
	>6 hrs			43 (72.9)		

High consumption of pork, fish, meat, oil and less consumption of fruits, vegetables and water are significantly associated with TNBC. Smoked meat and smoked vegetable consumption were not significant (Table 2).

Table 2. Food habits of TNBC patients

Food habits	**Categories**	**Frequency (%)**	**Chi square Value**	**P Value**
Pork	Never	10 (16.9)	25.78	**0.000**
	Yes	49 (83.1)		
Fish	Never	8 (13.6)	31.339	**0.000**
	Yes	51 (86.4)		
Chicken	Never	2 (3.4)	51.271	**0.000**
	Yes	57 (96.6)		
Mutton	Never	52 (88.1)	34.322	**0.000**
	Yes	7 (11.9)		
Beef	Never	41 (69.5)	8.966	**0.003**
	Yes	18 (30.5)		
Fruits	No	12 (20.3)	20.763	**0.000**
	Yes	47 (79.7)		
Vegetables	Less	9 (15.3)	28.492	**0.000**
	Regularly	50 (84.7)		
Saum	No	29 (49.2)	0.017	0.896
	Yes	30 (50.8)		
Smoked meat	No	32 (54.2)	0.424	0.515
	Yes	27 (45.8)		
Smoked vegetables	No	37 (62.7)	3.814	0.051
	Yes	22 (37.3)		

Salt intake	Normal	40 (67.8)	7.475	**0.006**
	Heavy	19 (32.2)		
Oil intake	Normal	50 (84.7)	28.492	**0.000**
	Heavy	9 (15.3)		
Water intake/day	<2 L	52 (88.1)	34.322	**0.000**
	>2 L	7 (11.9)		

Table 3. Lifestyle habits of TNBC patients

Lifestyle Habits	**Categories**	**Frequency (%)**	**Chi square Value**	**P Value**
Khuva/ Bettlenut	No	21 (35.6)	4.898	0.027
	Yes	38 (64.4)		
Gutkha	No	58 (98.3)	55.068	0.000
	Yes	1 (1.7)		
Sahdah	No	18 (30.5)	8.966	0.003
	Yes	41 (69.5)		
Khaini	No	56 (94.9)	47.61	0.000
	Yes	3 (5.1)		
Tuibur	No	46 (78.0)	18.458	0.000
	Yes	13 (22.0)		
Smoking	No	38 (64.4)	4.898	0.027
	Yes	21 (35.6)		
Alcohol	No	57 (96.6)	51.271	0.000
	Yes	2 (3.4)		

Consumption of tobacco products/ alcohol were highly significant with TNBC (Table 3). Different smoke, liquid form (tuibur) and smoke-less (Sahdah and khaini) forms of tobacco were practiced by the cases. Alcohol consumption also included the branded and local varieties.

Table 4. Biological and health status of TNBC patients				
Biological and health status	**Categories**	**Frequency (%)**	**Chi square Value**	**P Value**
Age at menarche	<13	8 (13.6)	51.271	**0.000**
	>13	51 (86.4)		
No of children	No child	11 (18.6)	31.339	**0.000**
	< 3 Child	27 (45.8)		
	>3 child	21 (35.6)		
Age at first delivery	Not delivered	11 (18.6)	6.644	**0.036**
	<20 years	15 (25.4)		
	21 -30 years	31 (52.5)		
	>30 years	2 (3.4)		
Breast feeding	No	11 (18.6)	29.881	**0.000**
	Yes	48 (81.4)		
Total Duration breast feeding	< 7 years	42 (71.2)	23.203	**0.000**
	>7 years	17 (28.8)		
Miscarriage	No	43 (72.9)	10.593	**0.001**
	Yes	16 (27.1)		
Menopause	< 45 years	21 (35.6)	12.356	0.000
	> 45 years	38 (64.4)		
Any type of cancer besides breast cancer	No	57 (96.6)	4.898	0.027
	Yes	2 (3.4)		
Co morbidities	No	39 (66.1)	51.271	0.000
	Yes	20 (33.9)		
First degree relatives	No	57 (96.6)	6.119	0.013
	Yes	2 (3.4)		
Second degree relatives	No	47 (79.7)	51.271	0.000
	Yes	12 (20.3)		

First second degree relatives Ovarian Cancer?	No	56 (94.9)	20.763	0.000
	Yes	3 (5.1)		
First second degree relatives other type cancer	No	15 (25.4)	47.61	0.000
	Yes	44 (74.6)		

Menarche, parity/ breast feeding/ miscarriage/ late menopausal age / familial cancer/ co-morbidities (diabetes and hypertension) / blood relatives with cancers are significantly associated with TNBC (Table 4). These findings are made under the category of reproductive history.

Table 5. Clinical Parameters of TNBC patients

Clinical parameters	**Categories**	**Frequency (%)**	**Chi square value**	**P Value**
Breast side	Left	33 (55.9)	25.78	**0.000**
	Right	25 (42.4)		
	Both	1 (1.7)		
Fibroadenoma	No	49 (83.1)	67.25	**0.000**
	Yes	10 (16.9)		
Grossed tumour size	Upto 2.5 cm	33 (55.9)	17.22	**0.000**
	2.6-5 cm	19 (32.2)		
	> 5 cm	7 (11.9)		
Location	Upper inner quadrant	11 (18.6)	25.71	**0.000**
	Upper outer quadrant	22 (37.3)		
	Lower inner quadrant	2 (3.4)		
	Lower outer quadrant	12 (20.3)		
	Central quadrant	8 (13.6)		
	All Quadrant	4 (6.8)		
Impression	IDC	49 (83.1)	8.32	**0.040**
	ILC	1 (1.7)		
	Medullary	9 (15.3)		
Grade	No Grade	10 (16.9)		
	Grade I	11 (18.6)		
	Grade II	24 (40.7)	10.59	**0.001**

	Grade III	14 (23.7)		
Lymph nodes	Negative	42 (71.2)	10.59	**0.001**
	Positive	17 (28.8)		
No of Lymph nodes	0/6 - 0/49	42 (71.2)	18.49	**0.000**
	1/13 - 21/28	17 (28.8)		
DCIS	Absent	46 (78.0)		
	Present	13 (22.0)	25.78	**0.000**
Recurrence	No	49 (83.1)		
	Yes	10 (16.9)		
Alive or deceased	Dead	9 (15.3)	28.49	**0.000**
	Alive	50 (84.7)		

IDC, tumor grade, Lymph node metastasis and DCIS are significantly associated with TNBC (Table 5).

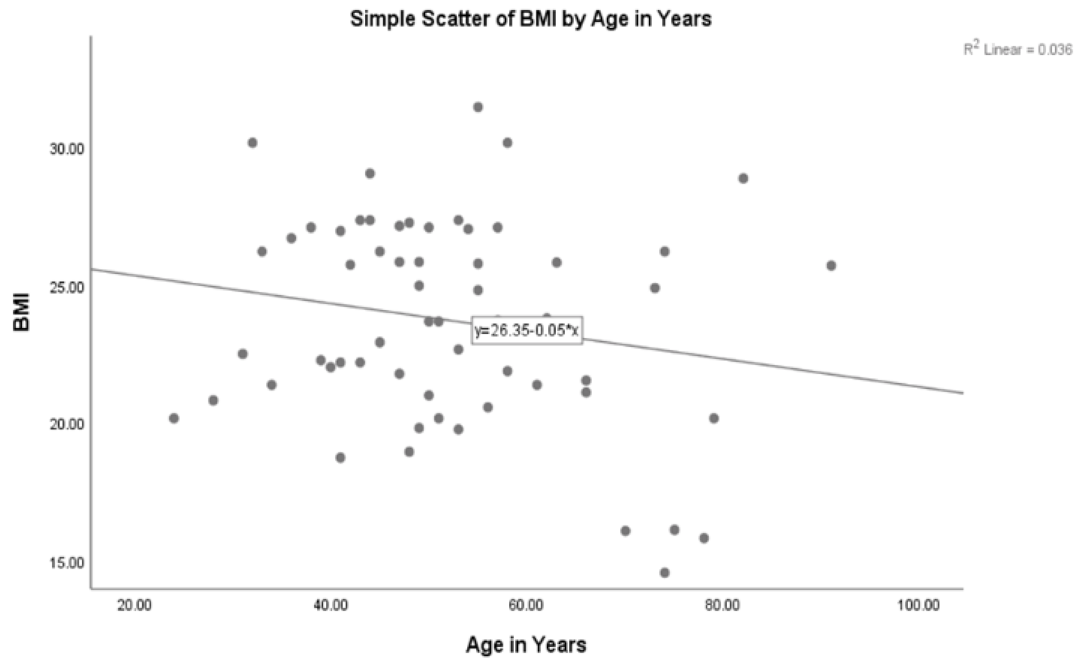


Figure 6: BMI of TNBC patients.

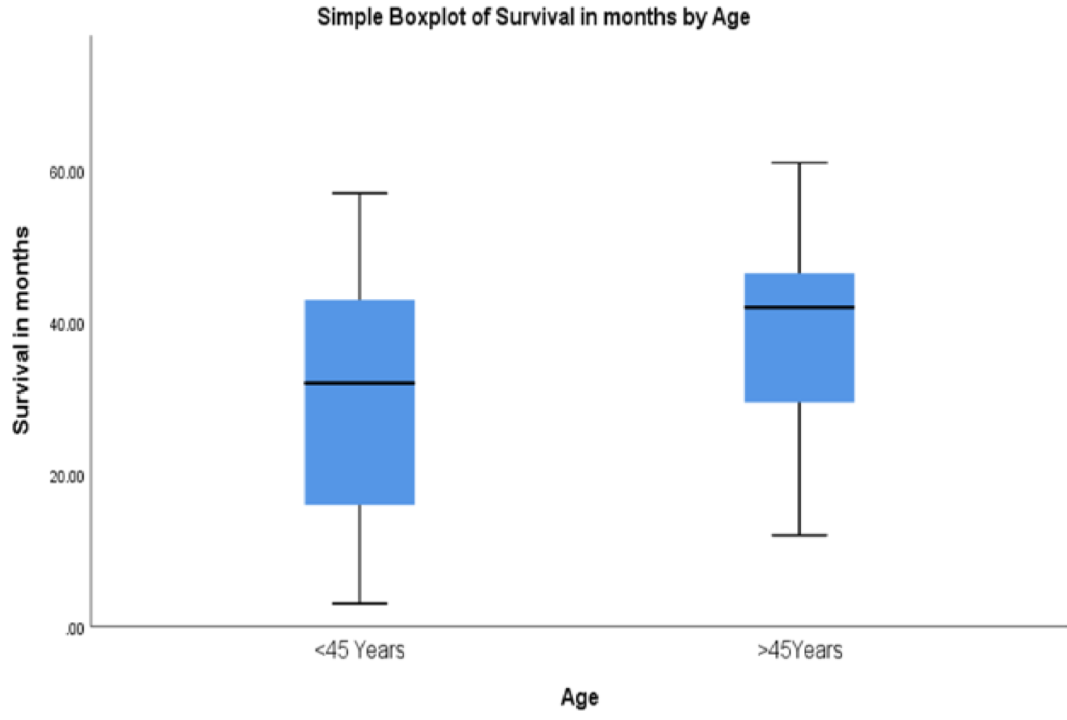


Figure 7: Overall survival of TNBC patients.

BMI was found to be higher in younger age TNBC patients (Figure 6). Overall survival was higher in older age (> 45 years) TNBC patients (Figure 7).

Table 6. Association between Demographic Variables and food habits Versus the recurrence of TNBC

Demographic Variables and food habits	Categories	Recurrence (%)		Chi square value	P Value
		No	Yes		
Age	<45 Years	30.6	50.0	1.393	0.238
	>45Years	69.4	50.0		
How often do you take exercise?	Never	65.3	80.0	0.825	0.662
	Once a week	16.3	10.0		
	Everyday	18.4	10.0		
	1-3 hrs	4.1	0.0	0.621	0.733
How many hours of sleep do you get?	3-6 hrs	22.4	30.0		
	>6 hrs	73.5	70.0		
Pork	Never	16.3	20.0	0.080	0.778
	Yes	83.7	80.0		
Fish	Never	10.2	30.0	2.777	0.096

	Yes	89.8	70.0		
Chicken	Never	4.1	0.0	0.422	0.516
	Yes	95.9	100.0		
Mutton	Never	87.8	90.0	0.040	0.841
	Yes	12.2	10.0		
Beef	Never	69.4	70.0	0.001	0.969
	Yes	30.6	30.0		
Fruits	No	22.4	10.0	0.794	0.373
	Yes	77.6	90.0		
Vegetables	Less	14.3	20.0	0.210	0.647
	Regularly	85.7	80.0		
Saum	No	46.9	60.0	0.567	0.451
	Yes	53.1	40.0		
Smoked meat	No	53.1	60.0	0.161	0.688
	Yes	46.9	40.0		
Smoked vegetables	No	61.2	70.0	0.274	0.601
	Yes	38.8	30.0		
Salt intake	Normal	73.5	40.0	4.261	0.039*
	Heavy	26.5	60.0		
Oil intake	Normal	85.7	80.0	0.210	0.647
	Heavy	14.3	20.0		
Water intake/day	<2 L	87.8	90.0	0.040	0.841
	>2 L	12.2	10.0		
Kuhva/Bettlenut	No	36.7	30.0	0.164	0.685
	Yes	63.3	70.0		
Gutkha	No	100.0	90.0	4.984	0.026*
	Yes	0.0	10.0		
Sahdah	No	30.6	30.0	0.001	0.969
	Yes	69.4	70.0		
Khaini	No	93.9	100.0	0.645	0.422
	Yes	6.1	0.0		
Tuibur	No	83.7	50.0	5.482	0.019*
	Yes	16.3	50.0		
Smoking	No	67.3	50.0	1.090	0.296
	Yes	32.7	50.0		
Alcohol	No	95.9	100.0	0.422	0.516
	Yes	4.1	0.0		

Consumption of Salt, Gutkha and tuibur are significantly associated with TNBC recurrence in our patient cohort (Table 6). The biological parameters are not significantly associated with TNBC recurrence in our patient cohort (Table 7).

Table 7: Association between biological parameters of TNBC patients with recurrence

Biological parameters	**Categories**	**Recurrence (%)**		**Chi square value**	**P Value**
		No	Yes		
Age at menarche	<13	14.3	10.0	0.130	0.718
	>13	85.7	90.0		
No of children	No child	20.4	10.0	1.122	0.571
	< 3 Child	42.9	60.0		
	>3 child	36.7	30.0		
Age at first delivery	Not delivered	20.4	10.0	1.728	0.631
	<20 years	26.5	20.0		
	21 -30 years	49.0	70.0		
	>30 years	4.1	0.0		
Breast feeding	No	20.4	10.0	0.593	0.441
	Yes	79.6	90.0		
Total Duration breast feeding	< 7 years	69.4	80.0	0.456	0.499
	<7 years	30.6	20.0		
Miscarriage	No	71.4	80.0	0.309	0.578
	Yes	28.6	20.0		
Menopause	< 45 years	32.7	50.0	1.090	0.296
	> 45 years	67.3	50.0		

Table 8: Association between socio-demographic variables and biological parameters of TNBC patients with other types of cancer

Demographic Variables and Food habits items	Categories	Any other cancer			
		No	Yes	Chi square value	P Value
		N (%)	(N) %		
Age	<45 Years	35.1	0.0	1.062	0.303
	>45Years	64.9	100.0		
How often do you take exercise?	Never	66.7	100.0	0.983	0.612
	Once a week	15.8	0.0		
	Everyday	17.5	0.0		
How many hours of sleep do you get?	1-3 hrs	3.5	0.0	0.770	0.68
	3-6 hrs	24.6	0.0		
	>6 hrs	71.9	100.0		
Pork	Never	15.8	50.0	1.606	0.205
	Yes	84.2	50.0		
Fish	Never	14.0	0.0	0.325	0.569
	Yes	86.0	100.0		
Chicken	Never	3.5	0.0	0.073	0.788
	Yes	96.5	100.0		
Mutton	Never	89.5	50.0	2.879	0.09
	Yes	10.5	50.0		
Beef	Never	68.4	100.0	0.909	0.34
	Yes	31.6	0.0		
Fruits	No	21.1	0.0	0.529	0.467
	Yes	78.9	100.0		
Vegetables	Less	14.0	50.0	1.933	0.164
	Regularly	86.0	50.0		
Saum	No	49.1	50.0	0.001	0.981
	Yes	50.9	50.0		

Smoked meat	No	54.4	50.0	0.015	0.903
	Yes	45.6	50.0		
Smoked vegetables	No	64.9	0.0	3.482	0.062
	Yes	35.1	100.0		
Salt intake	Normal	68.4	50.0	0.300	0.584
	Heavy	31.6	50.0		
Oil intake	Normal	84.2	100.0	0.373	0.542
	Heavy	15.8	0.0		
Water intake/day	<2 L	87.7	100.0	0.279	0.598
	>2 L	12.3	0.0		
Kuhva/Bettlenut	No	35.1	50.0	0.187	0.665
	Yes	64.9	50.0		
Gutkha	No	98.2	100.0	0.036	0.85
	Yes	1.8	0.0		
Sahdah	No	28.1	100.0	4.715	.030*
	Yes	71.9	0.0		
Khaini	No	94.7	100.0%	0.111	0.739
	Yes	5.3	0.0		
Tuibur	No	77.2	100.0	0.585	0.444
	Yes	22.8	0.0		
Smoking	No	63.2	100.0	1.144	0.285
	Yes	36.8	0.0		
Alcohol	No	96.5	100.0	0.073	0.788
	Yes	3.5	0.0		
Age at menarche	<13	12.3	50.0	2.345	0.126
	>13	87.7	50.0		
No of children	No child	19.3	0.0	0.515	0.773
	< 3 Child	45.6	50.0		
	>3 child	35.1	50.0		
Age at first delivery	Not	19.3	0.0	0.950	0.813

	delivered				
	<20 years	24.6	50.0		
	21 -30 years	52.6	50.0		
	>30 years	3.5	0.0		
Breast feeding	No	19.3	0.0	0.474	0.491
	Yes	80.7	100.0		
Total Duration breast feeding	< 7 years	71.9	50.0	0.453	0.501
	<7 years	28.1	50.0		
Miscarriage	No	73.7	50.0	0.548	0.459
	Yes	26.3	50.0		
Menopause	< 45 years	36.8	0.0	1.144	0.285
	> 45 years	63.2	100.0		
Co-morbidities	No	68.4	0.0	4.037	0.045,*
	Yes	31.6	100.0		
First degree	No	96.5	100.0	0.073	0.788
	Yes	3.5	0.0		
Second degree	No	80.7	50.0	1.124	0.289
	Yes	19.3	50.0		
First second degree Ovarian Cancer?	No	96.5	50.0	8.653	0.003,*
	Yes	3.5	50.0		
First second degree other type cancer	No	24.6	50.0	0.659	0.417
	Yes	75.4	50.0		
Familial Nonfamilial	No	24.6	50.0	0.659	0.417
	Yes	75.4	50.0		

Other cancer types are not significantly associated with the above TNBC socio-demographic and biological parameters in our patient cohort. Other cancer types are significantly associated with sahdah, co-morbidities and relatives with ovarian cancer (Table 8).

Table 9. Association between socio demographic variables and biological parameters of TNBC patients with Co-morbidities.

Demographic Variables and Food habits items	Categories	Co-morbidities			
		No (N %)	Yes (N %)	Chi square value	P Value
Age	<45 Years	46.2	10.0	7.712	0.005*
	>45Years	53.8	90.0		
How often do you take exercise?	Never	71.8	60.0	3.859	0.145
	Once a week	17.9	10.0		
	Everyday	10.3	30.0		
How many hours of sleep do you get?	1-3 hrs	2.6	5.0	0.286	0.867
	3-6 hrs	23.1	25.0		
	>6 hrs	74.4	70.0		
Pork	Never	15.4	20.0	0.200	0.655
	Yes	84.6	80.0		
Fish	Never	10.3	20.0	1.071	0.301
	Yes	89.7	80.0		
Chicken	Never	0.0	10.0	4.037	0.045*
	Yes	100.0	90.0		
Mutton	Never	89.7	85.0	0.284	0.594
	Yes	10.3	15.0		
Beef	Never	79.5	50.0	5.422	0.020*
	Yes	20.5	50.0		
Fruits	No	20.5	20.0	0.002	0.963
	Yes	79.5	80.0		
Vegetables	Less	15.4	15.0	0.002	0.969
	Regularly	84.6	85.0		

Saum	No	56.4	35.0	2.425	0.119
	Yes	43.6	65.0		
Smoked meat	No	59.0	45.0	1.040	0.308
	Yes	41.0	55.0		
Smoked vegetables	No	69.2	50.0	2.091	0.148
	Yes	30.8	50.0		
Salt intake	Normal	71.8	60.0	0.842	0.359
	Heavy	28.2	40.0		
Oil intake	Normal	89.7	75.0	2.223	0.136
	Heavy	10.3	25.0		
Water intake/day	<2 L	92.3	80.0	1.915	0.166
	>2 L	7.7	20.0		
Kuhva/Bettlenut	No	28.2	50.0	2.739	0.098
	Yes	71.8	50.0		
Gutkha	No	97.4	100.0	0.522	0.47
	Yes	2.6	0.0		
Sahdah	No	25.6	40.0	1.286	0.257
	Yes	74.4	60.0		
Khaini	No	92.3	100.0	1.621	0.203
	Yes	7.7	0.0		
Tuibur	No	74.4	85.0	0.871	0.351
	Yes	25.6	15.0		
Smoking	No	69.2	55.0	1.168	0.280
	Yes	30.8	45.0		
Alcohol	No	94.9	100.0	1.062	0.303
	Yes	5.1	0.0		
Age at menarche	<13	15.4	10.0	0.327	0.567
	>13	84.6	90.0		
No of children	No child	25.6	5.0	6.442	0.040*
	< 3 Child	48.7	40.0		

	>3 child	25.6	55.0		
Age at first delivery	Not delivered	25.6	5.0	8.141	0.043*
	<20 years	15.4	45.0		
	21 -30 years	56.4	45.0		
	>30 years	2.6	5.0		
Breast feeding	No	23.1	10.0	1.490	0.222
	Yes	76.9	90.0		
Total Duration breast feeding	< 7 years	79.5	55.0	3.865	0.049*
	>7 years	20.5	45.0		
Miscarriage	No	74.4	70.0	0.127	0.721
	Yes	25.6	30.0		
Menopause	< 45 years	46.2	15.0	5.597	0.018*
	> 45 years	53.8	85.0		
First degree	No	97.4	95.0	0.240	0.625
	Yes	2.6	5.0		
Second degree	No	79.5	80.0	0.002	0.963
	Yes	20.5	20.0		
First second degree Ovarian Cancer?	No	94.9	95.0	0.000	0.983
	Yes	5.1	5.0		
First second degree other type cancer	No	30.8	15.0	1.734	0.188
	Yes	69.2	85.0		
Familial_Nonfamilial	No	30.8	15.0	1.734	0.188
	Yes	69.2	85.0		

Comorbidities are not significantly associated with the above TNBC socio-demographic parameters in our patient cohort. Comorbidities are significantly associated with no. of children, age at first delivery, breast feeding duration and menopause (Table 9).

Table 10: Association between socio demographic variables and biological parameters of TNBC patients with Familial_Nonfamilial occurrence

Demographic Variables and Food habits items	Categories	Familial_Nonfamilial			
		No (N %)	Yes (N %)	Chi square value	P Value
Age	<45 Years	26.7	36.4	0.469	0.493
	>45Years	73.3	63.6		
How often do you take exercise?	Never	73.3	65.9	1.174	0.556
	Once a week	6.7	18.2		
	Everyday	20.0	15.9		
How many hours of sleep do you get?	1-3 hrs	0.0	4.5	0.940	0.625
	3-6 hrs	20.0	25.0		
	>6 hrs	80.0	70.5		
Pork	Never	6.7	20.5	1.511	0.219
	Yes	93.3	79.5		
Fish	Never	13.3	13.6	0.001	0.976
	Yes	86.7	86.4		
Chicken	Never	6.7	2.3	0.659	0.417
	Yes	93.3	97.7		
Mutton	Never	86.7	88.6	0.042	0.839
	Yes	13.3	11.4		
Beef	Never	80.0	65.9	1.048	0.306
	Yes	20.0	34.1		
Fruits	No	13.3	22.7	0.609	0.435
	Yes	86.7	77.3		
Vegetables	Less	20.0	13.6	0.350	0.554
	Regularly	80.0	86.4		
Saum	No	60.0	45.5	0.947	0.330
	Yes	40.0	54.5		

Smoked meat	No	60.0	52.3	0.269	0.604
	Yes	40.0	47.7		
Smoked vegetables	No	46.7	68.2	2.214	0.137
	Yes	53.3	31.8		
Salt intake	Normal	80.0	63.6	1.372	0.241
	Heavy	20.0	36.4		
Oil intake	Normal	93.3	81.8	1.147	0.284
	Heavy	6.7	18.2		
Water intake/day	<2 L	100.0	84.1	2.708	0.1
	>2 L	0.0	15.9		
Kuhva/Bettlenut	No	60.0	27.3	5.227	.022*
	Yes	40.0	72.7		
Gutkha	No	100.0	97.7	0.347	0.556
	Yes	0.0	2.3		
Sahdah	No	33.3	29.5	0.076	0.783
	Yes	66.7	70.5		
Khaini	No	100.0	93.2	1.078	0.299
	Yes	0.0	6.8		
Tuibur	No	86.7	75.0	0.886	0.346
	Yes	13.3	25.0		
Smoking	No	60.0	65.9	0.170	0.680
	Yes	40.0	34.1		
Alcohol	No	100.0	95.5	0.706	0.401
	Yes	0.0	4.5		
Age at menarche	<13	13.3	13.6	0.001	0.976
	>13	86.7	86.4		
No of children	No child	20.0	18.2	0.052	0.974
	< 3 Child	46.7	45.5		
	>3 child	33.3	36.4		
Age at first	Not	20.0	18.2	1.329	0.722

delivery	delivered				
	<20 years	33.3	22.7		
	21 -30 years	46.7	54.5		
	>30 years	0.0	4.5		
Breast feeding	No	20.0	18.2	0.024	0.876
	Yes	80.0	81.8		
Total Duration breast feeding	< 7 years	66.7	72.7	0.200	0.654
	<7 years	33.3	27.3		
Miscarriage	No	73.3	72.7	0.002	0.964
	Yes	26.7	27.3		
Menopause	< 45 years	33.3	36.4	0.045	0.832
	> 45 years	66.7	63.6		
First degree	No	100.0	95.5	0.706	0.401
	Yes	0.0	4.5		
Second degree	No	93.3	75.0	2.321	0.128
	Yes	6.7	25.0		
First second degree Ovarian Cancer?	No	93.3	95.5	0.104	0.747
	Yes	6.7	4.5		
First second degree other type cancer	No	100.0	0.0	59.000	0.000,*
	Yes	0.0	100.0		

Familial cancers are significantly associated with bettlenut consumption. Familial cancers are significantly associated with first or second degree relatives with other types of cancer (Table 10).

Table 11: Association between socio demographic variables and biological parameters of TNBC patients with Tumour type

Demographic Variables and Food habits items	Categories	Tumour Type				
		IDC (N %)	ILC (N %)	Medullary (N %)	Chi square value	P Value
Age	<45 Years	36.7	0.0	22.2	1.236	0.539
	>45Years	63.3	100.0	77.8		
How often do you take exercise?	Never	69.4	0.0	66.7	8.669	0.070
	Once a week	18.4	0.0	0.0		
	Everyday	12.2	100.0	33.3		
How many hours of sleep do you get?	1-3 hrs	2.0	0.0	11.1	5.708	0.222
	3-6 hrs	24.5	100.0	11.1		
	>6 hrs	73.5	0.0	77.8		
Pork	Never	16.3	0.0	22.2	0.395	0.821
	Yes	83.7	100.0	77.8		
Fish	Never	14.3	0.0	11.1	0.225	0.894
	Yes	85.7	100.0	88.9		
Chicken	Never	4.1	0.0	0.0	0.422	0.810
	Yes	95.9	100.0	100.0		
Mutton	Never	85.7	100.0	100.0	1.621	0.445
	Yes	14.3	0.0	0.0		
Beef	Never	67.3	0.0	88.9	3.981	0.137
	Yes	32.7	100.0	11.1		
Fruits	No	18.4	0.0	33.3	1.311	0.519
	Yes	81.6	100.0	66.7		
Vegetables	Less	12.2	0.0	33.3	2.799	0.247
	Regularly	87.8	100.0	66.7		
Saum	No	44.9	0.0	77.8	4.272	0.118
	Yes	55.1	100.0	22.2		

Smoked meat	No	53.1	100.0	55.6	0.877	0.645
	Yes	46.9	0.0	44.4		
Smoked vegetables	No	59.2	100.0	77.8	1.729	0.421
	Yes	40.8	0.0	22.2		
Salt intake	Normal	65.3	0.0	88.9	4.078	0.130
	Heavy	34.7	100.0	11.1		
Oil intake	Normal	83.7	100.0	88.9	0.343	0.842
	Heavy	16.3	0.0	11.1		
Water intake/day	<2 L	87.8	0.0	100.0	8.647	0.013*
	>2 L	12.2	100.0	0.0		
Kuhva/Bettlenut	No	38.8	0.0	22.2	1.471	0.479
	Yes	61.2	100.0	77.8		
Gutkha	No	98.0	100.0	100.0	0.208	0.901
	Yes	2.0	0.0	0.0		
Sahdah	No	28.6	100.0	33.3	2.398	0.301
	Yes	71.4	0.0	66.7		
Khaini	No	98.0	100.0	77.8	6.471	0.039*
	Yes	2.0	0.0	22.2		
Tuibur	No	81.6	0.0	66.7	4.591	0.101
	Yes	18.4	100.0	33.3		
Smoking	No	69.4	100.0	33.3	4.874	0.087
	Yes	30.6	0.0	66.7		
Alcohol	No	98.0	100.0	88.9	1.946	0.378
	Yes	2.0	0.0	11.1		
Age at menarche	<13	14.3	100.0	0.0	7.809	0.020*
	>13	85.7	0.0	100.0		
No of children	No child	16.3	100.0	22.2	7.079	0.132
	< 3 Child	51.0	0.0	22.2		
	>3 child	32.7	0.0	55.6		
Age at first delivery	Not	16.3	100.0	22.2	7.222	0.301

	delivered					
	<20 years	22.4	0.0	44.4		
	21 -30 years	57.1	0.0	33.3		
	>30 years	4.1	0.0	0.0		
Breast feeding	No	16.3	100.0	22.2	4.613	0.1
	Yes	83.7	0.0	77.8		
Total Duration breast feeding	< 7 years	77.6	100.0	33.3	7.660	0.022*
	<7 years	22.4	0.0	66.7		
Miscarriage	No	69.4	100.0	88.9	1.842	0.398
	Yes	30.6	0.0	11.1		
Menopause	< 45 years	38.8	0.0	22.2	1.471	0.479
	> 45 years	61.2	100.0	77.8		
Co morbidities	No	63.3	100.0	77.8	1.236	0.539
	Yes	36.7	0.0	22.2		
First degree	No	95.9	100.0	100.0	0.422	0.81
	Yes	4.1	0.0	0.0		
Second degree	No	77.6	100.0	88.9	0.863	0.65
	Yes	22.4	0.0	11.1		
First second degree Ovarian Cancer?	No	93.9	100.0	100.0	0.645	0.724
	Yes	6.1	0.0	0.0		
First second-degree other type cancer	No	24.5	0.0	33.3	0.660	0.719
	Yes	75.5	100.0	66.7		
Familial_Nonfamilial	No	24.5	0.0	33.3	0.660	0.719
	Yes	75.5	100.0	66.7		

Tumor type is significantly associated with water consumption in TNBC patients. Tumor type is significantly associated with khaini, age at menarche and breast-feeding duration in TNBC patients (Table 11).

Table 12: Association between socio demographic variables and biological parameters of TNBC patients with Lymph node occurrence.

Demographic Variables and Food habits items	Categories	Lymph nodes			
		Negative (N %)	Positive (N %)	Chi square value	P Value
Age	<45 Years	26.2	52.9	3.865	0.049*
	>45Years	73.8	47.1		
How often do you take exercise?	Never	66.7	70.6	0.225	0.894
	Once a week	16.7	11.8		
	Everyday	16.7	17.6		
How many hours of sleep do you get?	1-3 hrs	2.4	5.9	2.431	0.297
	3-6 hrs	19.0	35.3		
	>6 hrs	78.6	58.8		
Pork	Never	16.7	17.6	0.008	0.928
	Yes	83.3	82.4		
Fish	Never	14.3	11.8	0.066	0.798
	Yes	85.7	88.2		
Chicken	Never	2.4	5.9	0.453	0.501
	Yes	97.6	94.1		
Mutton	Never	88.1	88.2	0.000	0.988
	Yes	11.9	11.8		
Beef	Never	71.4	64.7	0.258	0.612
	Yes	28.6	35.3		
Fruits	No	23.8	11.8	1.084	0.298
	Yes	76.2	88.2		
Vegetables	Less	14.3	17.6	0.106	0.745
	Regularly	85.7	82.4		
Saum	No	47.6	52.9	0.137	0.711
	Yes	52.4	47.1		

Smoked meat	No	52.4	58.8	0.202	0.653
	Yes	47.6	41.2		
Smoked vegetables	No	66.7	52.9	0.975	0.323
	Yes	33.3	47.1		
Salt intake	Normal	66.7	70.6	0.085	0.770
	Heavy	33.3	29.4		
Oil intake	Normal	85.7	82.4	0.106	0.745
	Heavy	14.3	17.6		
Water intake/day	<2 L	90.5	82.4	0.764	0.382
	>2 L	9.5	17.6		
Kuhva/Bettlenut	No	35.7	35.3	0.001	0.976
	Yes	64.3	64.7		
Gutkha	No	100.0	94.1	2.513	0.113
	Yes	0.0	5.9		
Sahdah	No	31.0	29.4	0.014	0.907
	Yes	69.0	70.6		
Khaini	No	95.2	94.1	0.031	0.859
	Yes	4.8	5.9		
Tuibur	No	81.0	70.6	0.757	0.384
	Yes	19.0	29.4		
Smoking	No	64.3	64.7	0.001	0.976
	Yes	35.7	35.3		
Alcohol	No	95.2	100.0	0.838	0.36
	Yes	4.8	0.0		
Age at menarche	<13	11.9	17.6	0.340	0.56
	>13	88.1	82.4		
No of children	No child	19.0	17.6	0.538	0.764
	< 3 Child	42.9	52.9		
	>3 child	38.1	29.4		
Age at first delivery	Not delivered	19.0	17.6	7.012	0.072

	<20 years	33.3	5.9		
	21 -30 years	42.9	76.5		
	>30 years	4.8	0.0		
Breast feeding	No	19.0	17.6	0.016	0.9
	Yes	81.0	82.4		
Total Duration breast feeding	< 7 years	64.3	88.2	3.384	0.066
	<7 years	35.7	11.8		
Miscarriage	No	78.6	58.8	2.388	0.122
	Yes	21.4	41.2		
Menopause	< 45 years	31.0	47.1	1.369	0.242
	> 45 years	69.0	52.9		
Co morbidities	No	57.1	88.2	5.221	0.022*
	Yes	42.9	11.8		
First degree	No	95.2	100.0	0.838	0.36
	Yes	4.8	0.0		
Second degree	No	83.3	70.6	1.213	0.271
	Yes	16.7	29.4		
First second degree Ovarian Cancer?	No	97.6	88.2	2.208	0.137
	Yes	2.4	11.8		
First second degree other type cancer	No	21.4	35.3	1.227	0.268
	Yes	78.6	64.7		
Familial_Nonfamilial	No	21.4	35.3	1.227	0.268
	Yes	78.6	64.7		

Lymph node metastasis is significantly associated with age and comorbidities in TNBC patients (Table 12). DCIS is significantly associated with Beef, fruits and sahdah consumption in TNBC patients (Table 13).

Table 13. Association between socio demographic variables and biological parameters of TNBC patients with DCIS occurrence.

Demographic Variables and Food habits	Categories	DCIS			
		Absent (N %)	Present (N %)	Chi square value	P Value
Age	<45 Years	37.0	23.1	0.871	0.351
	>45Years	63.0	76.9		
How often do you take exercise?	Never	63.0	84.6	2.164	0.339
	Once a week	17.4	7.7		
	Everyday	19.6	7.7		
How many hours of sleep do you get?	1-3 hrs	2.2	7.7	3.265	0.195
	3-6 hrs	19.6	38.5		
	>6 hrs	78.3	53.8		
Pork	Never	17.4	15.4	0.029	0.865
	Yes	82.6	84.6		
Fish	Never	15.2	7.7	0.490	0.484
	Yes	84.8	92.3		
Chicken	Never	4.3	0.0	0.585	0.444
	Yes	95.7	100.0		
Mutton	Never	87.0	92.3	0.278	0.598
	Yes	13.0	7.7		
Beef	Never	63.0	92.3	4.094	0.043*
	Yes	37.0	7.7		
Fruits	No	13.0	46.2	6.858	0.009*
	Yes	87.0	53.8		
Vegetables	Less	15.2	15.4	0.000	0.988
	Regularly	84.8	84.6		
Saum	No	43.5	69.2	2.690	0.101
	Yes	56.5	30.8		

Smoked meat	No	52.2	61.5	0.358	0.550
	Yes	47.8	38.5		
Smoked vegetables	No	58.7	76.9	1.440	0.23
	Yes	41.3	23.1		
Salt intake	Normal	63.0	84.6	2.160	0.142
	Heavy	37.0	15.4		
Oil intake	Normal	84.8	84.6	0.000	0.988
	Heavy	15.2	15.4		
Water intake/day	<2 L	84.8	100.0	2.245	0.134
	>2 L	15.2	0.0		
Kuhva/Bettlenut	No	37.0	30.8	0.169	0.681
	Yes	63.0	69.2		
Gutkha	No	97.8	100.0	0.287	0.592
	Yes	2.2	0.0		
Sahdah	No	37.0	7.7	4.094	0.043*
	Yes	63.0	92.3		
Khaini	No	95.7	92.3	0.235	0.628
	Yes	4.3	7.7		
Tuibur	No	76.1	84.6	0.429	0.512
	Yes	23.9	15.4		
Smoking	No	63.0	69.2	0.169	0.681
	Yes	37.0	30.8		
Alcohol	No	95.7	100.0	0.585	0.444
	Yes	4.3	0.0		
Age at menarche	<13	10.9	23.1	1.289	0.256
	>13	89.1	76.9		
No of children	No child	21.7	7.7	1.596	0.45
	< 3 Child	45.7	46.2		
	>3 child	32.6	46.2		
Age at first delivery	Not	21.7	7.7	4.185	0.242

	delivered				
	<20 years	28.3	15.4		
	21 -30 years	45.7	76.9		
	>30 years	4.3	0.0		
Breast feeding	No	21.7	7.7	1.318	0.251
	Yes	78.3	92.3		
Total Duration breast feeding	< 7 years	69.6	76.9	0.268	0.605
	<7 years	30.4	23.1		
Miscarriage	No	76.1	61.5	1.085	0.297
	Yes	23.9	38.5		
Menopause	< 45 years	37.0	30.8	0.169	0.681
	> 45 years	63.0	69.2		
Co morbidities	No	63.0	76.9	0.871	0.351
	Yes	37.0	23.1		
First degree	No	95.7	100.0	0.585	0.444
	Yes	4.3	0.0		
Second degree	No	78.3	84.6	0.253	0.615
	Yes	21.7	15.4		
First second degree Ovarian Cancer?	No	93.5	100.0	0.893	0.345
	Yes	6.5	0.0		
First second degree other type cancer	No	23.9	30.8	0.251	0.616
	Yes	76.1	69.2		
Familial_Nonfamilial	No	23.9	30.8	0.251	0.616
	Yes	76.1	69.2		

Table 14. Association between socio demographic variables and biological parameters of TNBC patients with Survival.

Demographic Variables and Food habits	Categories	Alive or Deceased			
		Dead (N %)	Alive (N %)	Chi square value	P Value
Age	<45 Years	44.4	32.0	0.527	0.468
	>45Years	55.6	68.0		
How often do you take exercise?	Never	77.8	66.0	0.489	0.783
	Once a week	11.1	16.0		
	Everyday	11.1	18.0		
How many hours of sleep do you get?	1-3 hrs	0.0	4.0	0.405	0.817
	3-6 hrs	22.2	24.0		
	>6 hrs	77.8	72.0		
Pork	Never	0.0	20.0	2.167	0.141
	Yes	100.0	80.0		
Fish	Never	22.2	12.0	0.680	0.41
	Yes	77.8	88.0		
Chicken	Never	11.1	2.0	1.933	0.164
	Yes	88.9	98.0		
Mutton	Never	88.9	88.0	0.006	0.939
	Yes	11.1	12.0		
Beef	Never	77.8	68.0	0.344	0.558
	Yes	22.2	32.0		
Fruits	No	11.1	22.0	0.558	0.455
	Yes	88.9	78.0		
Vegetables	Less	11.1	16.0	0.141	0.707
	Regularly	88.9	84.0		
Saum	No	44.4	50.0	0.094	0.759
	Yes	55.6	50.0		

Smoked meat	No	33.3	58.0	1.870	0.172
	Yes	66.7	42.0		
Smoked vegetables	No	44.4	66.0	1.516	0.218
	Yes	55.6	34.0		
Salt intake	Normal	33.3	74.0	5.777	0.016*
	Heavy	66.7	26.0		
Oil intake	Normal	88.9	84.0	0.141	0.707
	Heavy	11.1	16.0		
Water intake/day	<2 L	77.8	90.0	1.090	0.297
	>2 L	22.2	10.0		
Kuhva/ Bettlenut	No	44.4	34.0	0.363	0.547
	Yes	55.6	66.0		
Gutkha	No	100.0	98.0	0.183	0.669
	Yes	0.0	2.0		
Sahdah	No	33.3	30.0	0.040	0.842
	Yes	66.7	70.0		
Khaini	No	100.0	94.0	0.569	0.451
	Yes	0.0	6.0		
Tuibur	No	77.8	78.0	0.000	0.988
	Yes	22.2	22.0		
Smoking	No	66.7	64.0	0.024	0.878
	Yes	33.3	36.0		
Alcohol	No	100.0	96.0	0.373	0.542
	Yes	0.0	4.0		
Age at menarche	<13	11.1	14.0	0.054	0.816
	>13	88.9	86.0		
No of children	No child	11.1	20.0	1.871	0.392
	< 3 Child	33.3	48.0		
	>3 child	55.6	32.0		
Age at first delivery	Not	11.1	20.0	2.826	0.419

	delivered				
	<20 years	11.1	28.0		
	21 -30 years	77.8	48.0		
	>30 years	0.0	4.0		
Breast feeding	No	11.1	20.0	0.397	0.528
	Yes	88.9	80.0		
Total Duration breast feeding	< 7 years	44.4	76.0	3.703	0.054
	<7 years	55.6	24.0		
Miscarriage	No	66.7	74.0	0.208	0.649
	Yes	33.3	26.0		
Menopause	< 45 years	55.6	32.0	1.846	0.174
	> 45 years	44.4	68.0		
Co morbidities	No	77.8	64.0	0.646	0.421
	Yes	22.2	36.0		
First degree	No	88.9	98.0	1.933	0.164
	Yes	11.1	2.0		
Second degree	No	88.9	78.0	0.558	0.455
	Yes	11.1	22.0		
First second degree Ovarian Cancer?	No	100.0	94.0	0.569	0.451
	Yes	0.0	6.0		
First second degree other type cancer	No	22.2	26.0	0.057	0.811
	Yes	77.8	74.0		
Familial_Nonfamilial	No	22.2	26.0	0.057	0.811
	Yes	77.8	74.0		

Salt consumption is significantly associated with survival in TNBC patients (Table 14). All other demographic and biological parameters were insignificant.

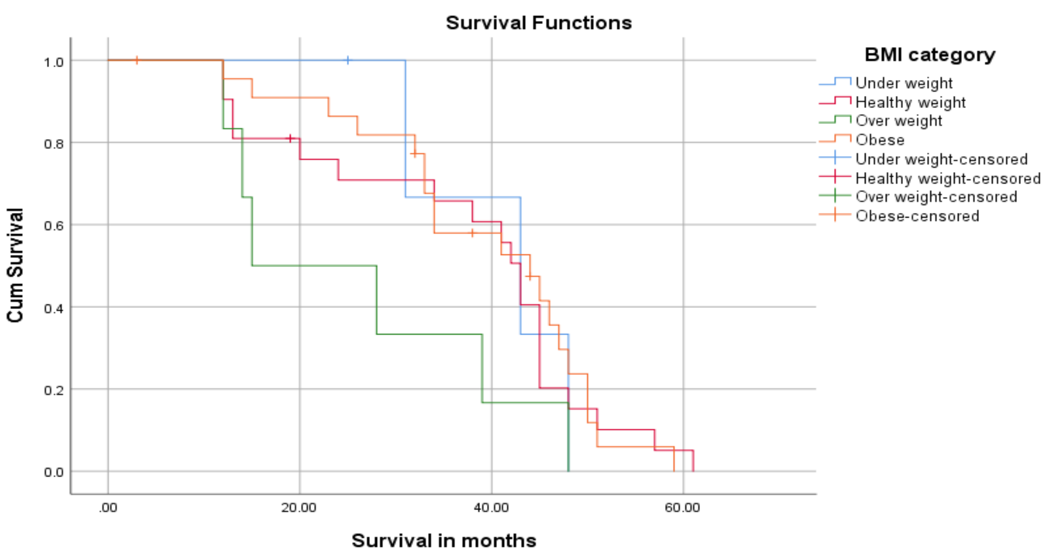


Figure 8. Overall survival in TNBC patients based on body weight.

Overall survival is significantly less in under-weight and over-weight in TNBC patients (Figure 8). For Asia-Pacific region, the BMI for normal is 18.5 to 22.9, Underweight is < 18.5 and Obese is ≥ 25.

Table 15: Association between socio demographic variables between Healthy controls and TNBC patients

Socio demographic variables	Categories	Group		OR (95% CI)	P Value
		Healthy N (%)	Cases N (%)		
Age	<45 Years	54 (46.2)	19 (35.2)	1.58 (0.81 to 3.07)	0.179
	>45Years	63 (53.8)	35 (64.8)		
How often do you take exercise?	No	61 (52.1)	36 (66.7)	0.545 (0.28 to 1.07)	0.076
	Yes	56 (47.8)	18(33.4)		
How many hours of sleep do you get?	Less	5 (4.3)	2 (3.7)	1.161 (0.22 to 6.19)	0.861
	Normal	112 (95.7)	52 (96.3)		

Exercise is mildly significant between the Healthy controls and TNBC patients (Table 15).

Table 16. Food habits between Healthy controls and TNBC patients.

Food habits	Categories	Group		OR (95% CI)	P Value
		Healthy N (%)	Case N (%)		
Pork	Never	14 (12.0)	9(16.7)	0.68 (0.27 to 1.68)	0.4043
	Yes	103 (88.0)	45(83.3)		
Fish	Never	9 (7.7)	8(14.8)	0.48 (0.17 to 1.32)	0.1546
	Yes	108 (92.3)	46(85.2)		
Chicken	Never	6 (5.1)	2(3.7)	1.40 (0.27 to 7.20)	0.6831
	Yes	111 (94.9)	52(96.3)		
Beef	Never	18 (15.4)	36(66.7)	0.09 (0.43 to 0.19)	< **0.0001**
	Yes	99 (84.6)	18(33.3)		
Fruits	No	3 (2.6)	11(20.4)	0.10 (0.03 to 0.38	**0.0008**
	Yes	114 (97.4)	43(79.6)		
Vegetables	Less	19 (16.2)	7(13.0)	1.30 (0.51 to 3.31)	0.5799
	Regularly	98 (83.8)	47(87.0)		
Saum	No	21 (17.9)	27(50.0)	0.22 (0.11 to 0.44)	< **0.0001**
	Yes	96 (82.1)	27(50.0)		
Smoked meat	No	26 (22.2)	30(55.6)	0.23 (0.11 to 0.45)	< **0.0001**
	Yes	91 (77.8)	24(44.4)		
Smoked vegetables	No	49 (41.9)	34(63.0)	0.42 (0.22 to 0.82)	**0.0112**
	Yes	68 (58.1)	20(37.0)		
Salt intake	Normal	88 (75.2)	38(70.4)	1.28 (0.62 to 2.62)	0.5043
	Heavy	29 (24.8)	16(29.6)		
Oil intake	Normal	86 (73.5)	45(83.3)	0.55 (0.24 to 1.26)	0.1617
	Heavy	31 (26.5)	9(16.7)		
Water intake/day	<2 L	54 (46.2)	47(87.0)	0.13 (0.05 to 0.30)	< **0.0001**
	>2 L	63 (53.8)	7(13.0)		

Table 17: Life style habits between Healthy controls and TNBC patients.

Lifestyle	Categories	Group		OR (95% CI)	P Value
		Healthy N (%)	Cases N (%)		
Kuhva/ Bettlenut	No	37(31.6)	19(35.2)	0.85 (0.43 to 1.68)	0.6448
	Yes	80(68.4)	35(64.8)		
Gutkha	No	100(85.5)	53(98.1)	0.11 (0.01 to 0.85)	**0.0350**
	Yes	17(14.5)	1(1.9)		
Sahdah	No	56(47.9)	17(31.5)	1.99 (1.01 to 3.94)	**0.0458**
	Yes	61(52.1)	37(68.5)		
Khaini	No	113(96.6)	51(94.4)	1.66 (0.35 t0 7.69)	0.5161
	Yes	4(3.4)	3(5.6)		
Tuibur	No	103(88.0)	42(77.8)	2.10 (0.89 to 4.92)	0.0869
	Yes	14(12.0)	12(22.2)		
Smoking	No	89(76.1)	36(66.7)	1.59 (0.78 to 3.22)	0.1993
	Yes	28(23.9)	18(33.3)		
Alcohol	No	117(100.0)	52(96.3)	11.19 (0.53 to 237.19)	0.1211
	Yes	0(0.0)	2(3.7)		

Regular consumption of Meat and smoked food were significant; Less intake of fruits and water are also significant between the Healthy controls and TNBC patients (Table 16). Consumption of Tobacco products is also significant between the Healthy controls and TNBC patients (Table 17). Breast feeding, Menopause and co-morbidities are significant between the Healthy controls and TNBC patients (Table 18).

Table 18: Health related parameters between Healthy controls and TNBC patients.

		Group			
Health related	**Categories**	**Healthy N (%)**	**Cases N (%)**	**OR (95% CI)**	**P Value**
Age at menarche	<13	28(23.9)	8(14.8)	1.80 (0.76 to 4.28)	0.1780
	>13	89(76.1)	46(85.2)		
No of children	No child	23(19.7)	11(20.4)	0.95 (0.42 to 2.14)	0.9136
	Child	94(80.3)	43(79.6)		
Age at first delivery	Not delivered	23(19.7)	11(20.4)	0.95 (0.42 to 2.14)	0.9136
	Delivered	94(80.3)	43(79.7)		
Breast feeding	No	24(20.5)	11(20.4)	1.01 (0.45 to 2.24)	0.9829
	Yes	93(79.5)	43(79.6)		
Total Duration breast feeding	< 7 years	102(87.2)	39(72.2)	2.61 (1.17 to 5.85)	**0.0193**
	<7 years	15(12.8)	15(27.8)		
Miscarriage	No	91(77.8)	39(72.2)	1.34 (0.64 to 2.81)	0.4298
	Yes	26(22.2)	15(27.8)		
Menopause	< 45 years	84(71.8)	19(35.2)	4.68 (2.35 to 9.33)	< **0.0001**
	> 45 years	33(28.2)	35(64.8)		
Any type of cancer besides breast cancer	No	117(100.0)	52(96.3)	10. 98 (0.51 to 232. 70)	0.1240
	Yes	0(0.0)	2(3.7)		
Co morbidities	No	117(100.0)	37(68.5)	109.6667 (6.43 to 1867.87)	**0.0012**
	Yes	0(0.0)	17(31.5)		
First degree	No	106(90.6)	52(96.3)	0.37 (0.07 to 1.73)	0.2073
	Yes	11(9.4)	2(3.7)		
Second degree	No	104(88.9)	43(79.6)	2.04 (0.85 to 4.92)	0.1099
	Yes	13(11.1)	11(20.4)		
First second degree Ovarian Cancer?	No	101(86.3)	51(94.4)	0.37 (0.10 to 1.33)	0.1288
	Yes	16(13.7)	3(5.6)		
First second degree other type cancer	No	45(38.5)	14(25.9)	1.78 (0.87 to 3.64)	0.1112
	Yes	72(61.5)	40(74.1)		

Explorative analysis of somatic variants

A total of 188 somatic SNVs and 647 genes were identified. The most common variants types were missense mutations (560), followed by nonsense mutations (39), frame shift mutations (37). Comprehensive details about these mutations can be referenced in Tables 19 and 20, while their visual representation is elucidated in Figure 9 A-B). Amino acid substitutions in proteins is mostly associated with genetic disorders, caused by single nucleotide mutations, known as missense mutations.

ID	Summary
Samples	15
Number of genes	647
Frame Shift Deletion	37
Frame Shift Insertion	6
In Frame Deletion	19
In Frame Insertion	3
Missense Mutation	560
Nonsense Mutation	39
Nonstop Mutation	1
Splice Site	20
Translation Start Site	3
Total	688

Table 19. Summary of somatic variants analysis.

The most common of these missense mutations is a single-nucleotide polymorphism (SNP) mutation, which was characterized by a base change and faced multiple genetic variants. Furthermore, the most prevalent single nucleotide variation (SNV) is C > T mutations (Figure 9C).

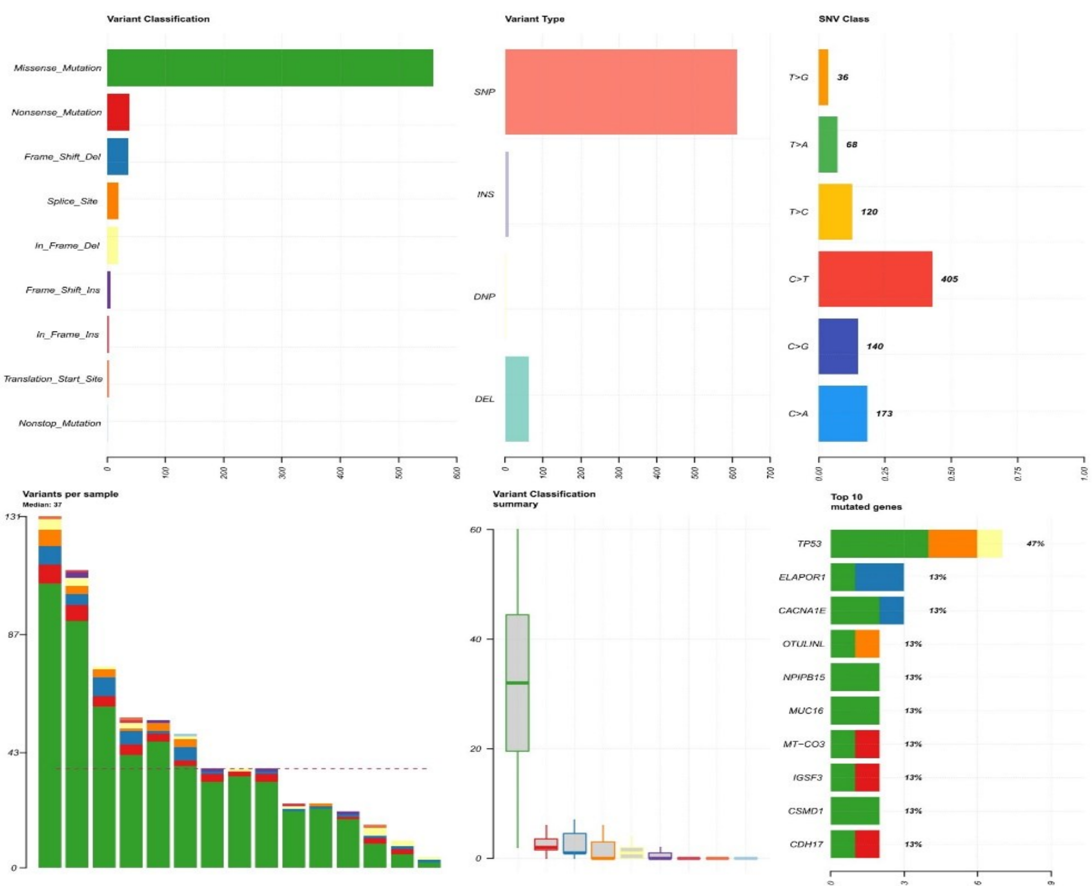


Figure 9. Summary of the filtered somatic variants in TNBC.

Panels A–C, illustrate the overall distribution of breast somatic variants while Panel D shows the number of variants in each sample. Panel E demonstrates variants per sample and Panel F shows the top 10 mutated genes.

The prevalence of C > T mutations suggests that most missense mutations are caused by cytosine (C) to thymine (T) changes at specific genomic loci, posing implications for understanding genetic mechanisms and disease associations. The number or variants per sample ranged from 4 to 131, with a median value of 37 (Figure 9D). Among the 10 most mutated genes, Maftools flags MUC16 as a gene that is usually found mutated in exome studies, therefore likely to be a passenger gene.

Table 20. Summary of somatic variants per samples.

Sample	Frame Shift Deletio	Frame Shift Insertions	In Frame Deletions	In Frame Insertions	Mis-sense Mutation	Non-sense Mutation	Non-stop Mutation	Splice Site	Translation Start Site	Total
T113	7	0	4	0	106	7	0	6	1	131
T188	4	2	3	1	92	6	0	3	0	111
T148	7	0	1	0	60	4	0	3	0	75
T89	5	0	2	1	42	4	0	1	1	56
T117	1	1	0	0	47	3	0	3	0	55
T86	5	0	1	0	38	2	1	3	0	50
T14	1	1	0	0	32	3	0	0	0	37
T146	0	0	1	0	34	2	0	0	0	37
T74	1	1	0	0	32	3	0	0	0	37
T34	1	0	1	1	21	0	0	0	0	24
T91	1	0	0	0	22	0	0	1	0	24
T157	1	1	0	0	18	1	0	0	0	21
T124	1	0	3	0	9	2	0	0	1	16
T147	1	0	2	0	5	2	0	0	0	10
T75	1	0	1	0	2	0	0	0	0	4

Frequently mutated genes in somatic variants

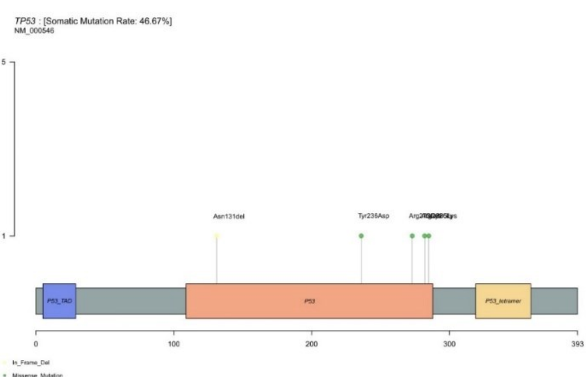
Several genes have been found to undergo mutations frequently in somatic TNBC. Among these, *TP53, CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD*, and *TLL1,* were the most frequently mutated genes, as illustrated in Figure 9F and Table 21. Moreover, *TP53* gene demonstrated the highest occurrence of genetic variants and short indels, present in 47% of the patients' tumour samples. Conversely, the remaining 15 genes collectively displayed a mutation frequency of 13%, as depicted in Figure 9F.

Table 21. Top 10 frequently mutated genes and the number of variants and altered samples.

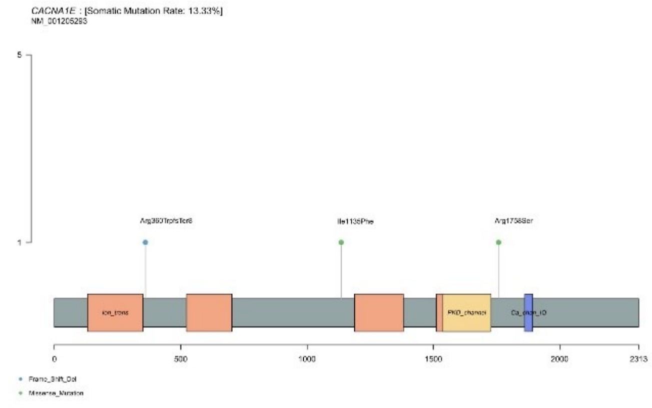
Genes	Frame Shift Del	FrameShift Ins	In Frame Del	In Frame Ins	Mis-sense Muta-tion	Non-sense Muta-tion	Splice Site	Total	Mutated Samples
TP53	0	0	1	0	4	0	2	7	7
CACNA1E	1	0	0	0	2	0	0	3	2
ELAPOR1	2	0	0	0	1	0	0	3	2
CDH17	0	0	0	0	1	1	0	2	2
CSMD1	0	0	0	0	2	0	0	2	2
IGSF3	0	0	0	0	1	1	0	2	2
MT-CO3	0	0	0	0	1	1	0	2	2
MUC16	0	0	0	0	2	0	0	2	2
NPIPB15	0	0	0	0	2	0	0	2	2
OTULINL	0	0	0	0	1	0	1	2	2

Furthermore, several oncogenes exhibit particular sites that undergo mutations more frequently than others, referred to as mutational hotspots. Visualization of these spots, along with other mutations, can be achieved through lollipop plots in Maftools. These plots offer a clear and impactful method to illustrate mutation locations on protein structures, based on protein domains obtained from the PFAM database (Figure 10 a - g).

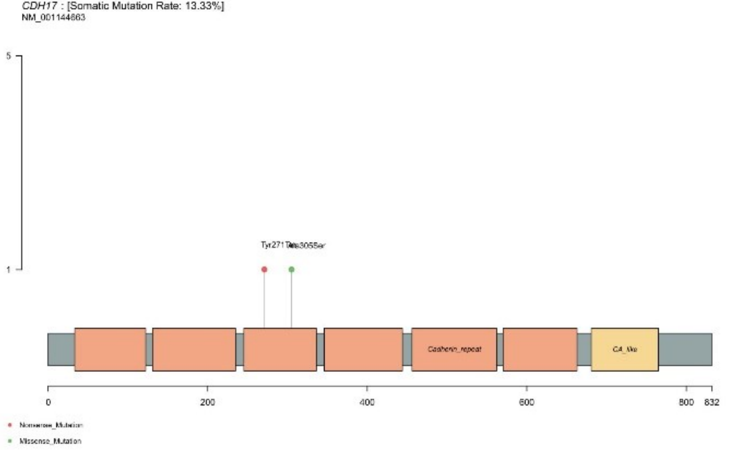
(a) ***TP53***



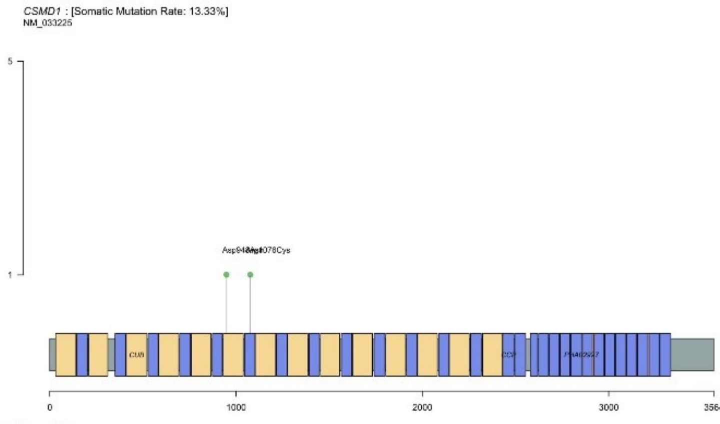
(b) ***CACNA1E***



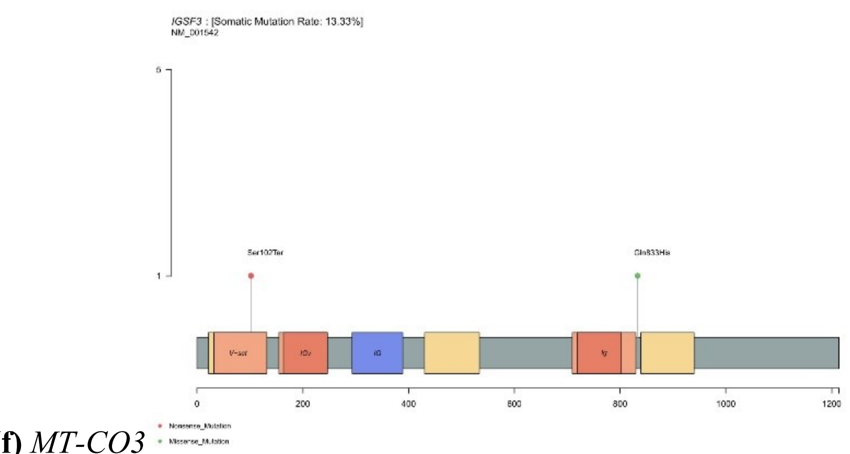
(c) ***CDH17***



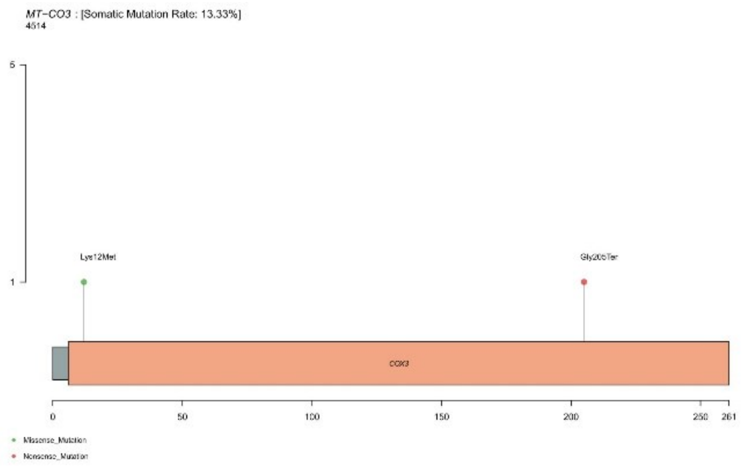
(d) *CSMD1*



(e) *JGSF3*



(f) *MT-CO3*



(g) *MUC16*

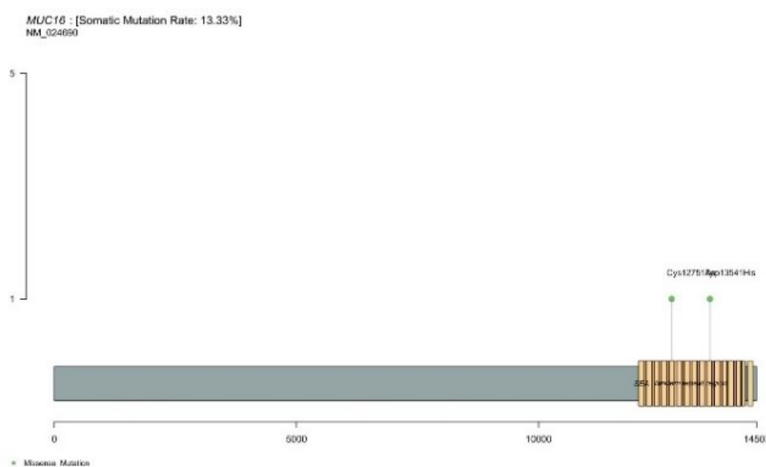


Figure 10. Top 10 frequently mutated genes showing protein mutation rate and the sites of variants

(a) *TP53,* (b) *CACNA1E,* (c) *CDH17,* (d) *CSMD1,* (e) *IGSF3,* (f) *MT-CO3,* (g) *MUC16*. Protein structures of three genes are not available in the pfam database and therefore only seven genes protein mutations are available.

Interacting gene pair analysis

A comprehensive somatic interaction analysis was conducted to investigate the relationship between the most commonly mutated genes within a cohort of patients with somatic breast conditions (Figure 11). The primary focus was on ascertaining whether these mutations tended to appear together or exhibited a pattern of mutual exclusivity. Two pairs of genes, *OTULINL* and *CDH17,* as well as *MUC16* and *MT-CO3*, displayed a pronounced tendency to co-occur. This observation was supported by statistical significance, with a p-value of 0.05 for each gene pair. This implies a non-random, potentially interdependent relationship between these gene combinations within the context of somatic breast conditions. These findings may indicate the potential interactions influencing the development or progression of somatic breast conditions, suggesting possible cooperative mechanism or shared pathways between these gene pairs. Further investigation into the functional

implications of these gene interactions could provide valuable insights for targeted therapeutic approaches or diagnostic strategies in breast pathology.

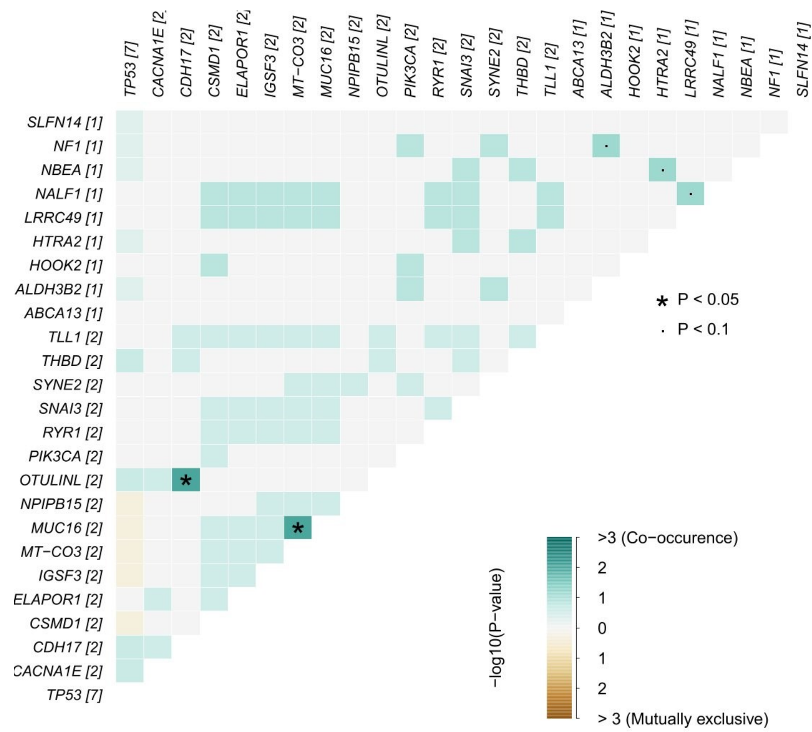


Figure 11. Interacting gene pair analysis of the 30 most mutated genes in somatic TNBC.

Somatic Mutational Signature Analysis

Mutational signature analysis revealed the presence of three distinct mutational signatures associated with *TP53* gene (Figure 12; Table 22). Accordingly, the signature 1 has a best match of COSMIC_6 mutational signature with an aetiology of defective DNA mismatch repair with a cosine similarity score of 0.811. Similarly, the signature 2 has a cosine similarity score of 0.771 with the mutational signature of COSMIC_6 also of aetiology of defective DNA mismatch repair (Figure 13). The defective DNA mismatch repair mechanism refers to a dysfunctional cellular machinery that is responsible for correcting errors in DNA replication. DNA mismatch repair (MMR) is responsible for maintaining the integrity of genetic

material in an organism. This fundamental function prevents mutations and genetic instability by identifying and repairing mismatches or small distortions that may occur during DNA replication.

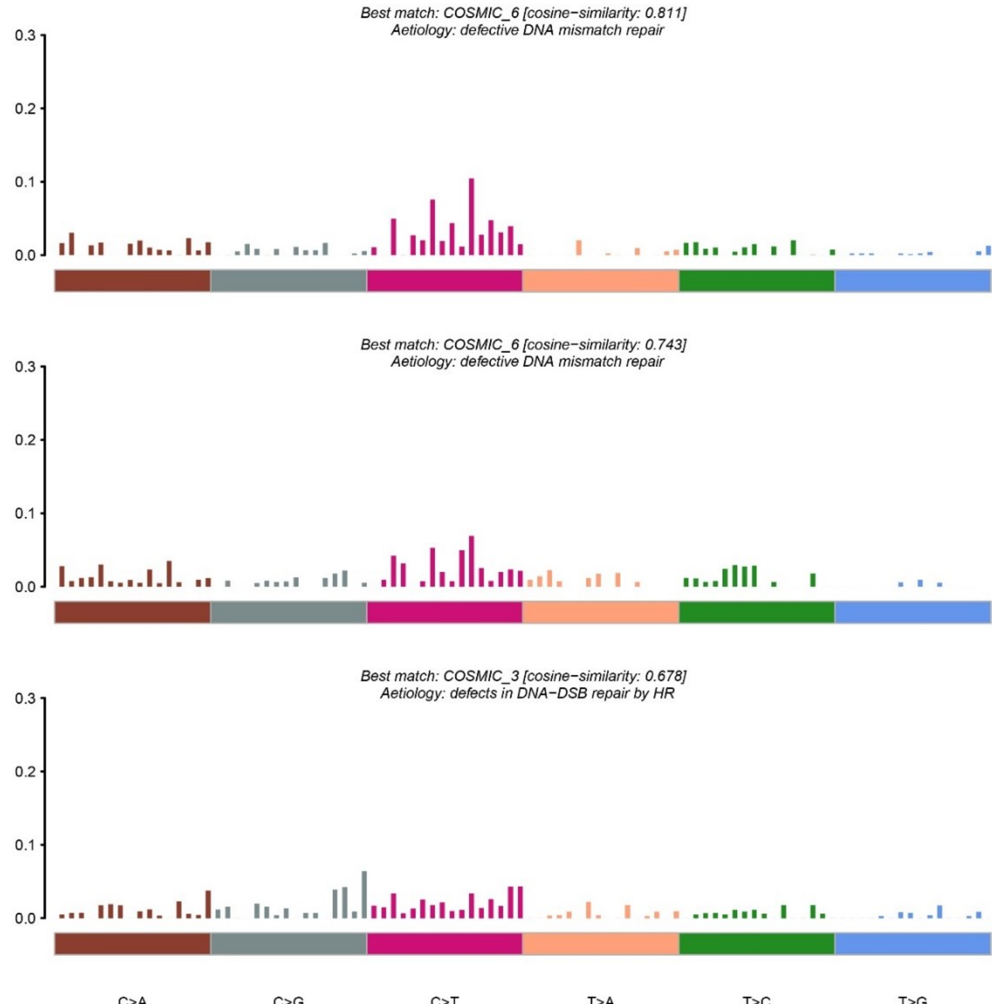


Figure 12. Mutational signature analysis with COSMIC database in somatic TNBC.

The signature 3 closely match the COSMIC_3 mutational signature, scoring 0.678 on the cosine similarity scale with an aetiology of defects in DNA−DSB repair by HR. Homologous recombination is a pathway to repair DNA double-strand breaks (DSBs). A number of defects or events can affect DNA DSB of HR, such as increased cancer and genetic diseases (Figure 14).

Table 22: Mutational signature analysis of somatic variants in TNBC.

Hugo Symbol	Group1	Group2	Number of mutated group1	Number of mutated group2	p_value	fdr
TP53	Signature_1	Rest	4 of 7	3 of 8	0.61927	0.92890
TP53	Signature_2	Rest	1 of 4	6 of 11	0.56923	0.92890
TP53	Signature_3	Rest	2 of 4	5 of 11	1	1

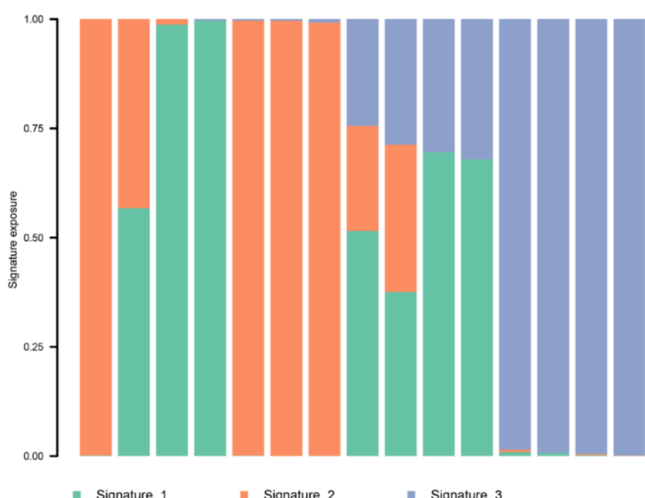


Figure 13. COSMIC mutational signatures across somatic TNBC samples.

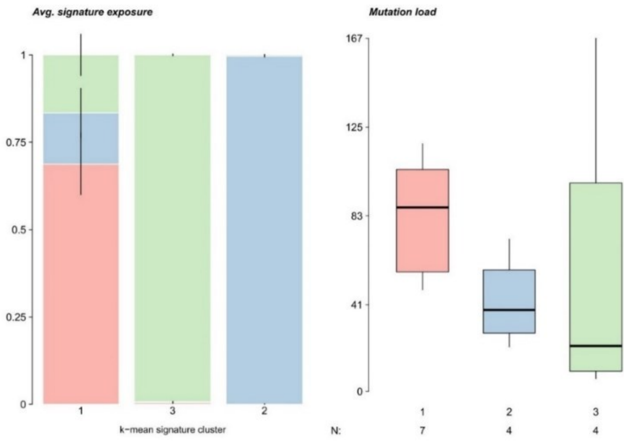


Figure 14. Average signature exposure and mutation load in somatic TNBC samples.

Pathway Analysis

A pathway analysis showed several genes associated with the oncogenic pathways (Figure 23). There are 22 genes associated with the PI3K-Akt signaling pathway. This pathway plays a critical role in regulating cell growth, survival, and various cellular processes. The MAPK signaling pathway, with 19 gene counts, is also of notable importance as it is involved in regulating cell proliferation and differentiation. Apoptosis, an essential mechanism for programmed cell death, had 13 genes associated with it. Furthermore, several pathways, such as the Rap1 signaling pathway, Phospholipase D signaling pathway, and Cellular senescence, had intermediate gene counts, suggesting their potential relevance in the study's biological context. In contrast, pathways with lower gene counts may still hold significance. For example, pathways like Starch and sucrose metabolism and Carbohydrate digestion and absorption had six gene counts, suggesting a potential role in metabolic processes.

Table 23: Oncogenic pathways associated with variants and their related somatic genes in TNBC.

EGG_PATHWAY	Genes	Count	%	P-Value
PI3K-Akt signaling pathway	*HSP90AA1, CSF3R, LAMA2, VWF, CSF1, LAMB2, TSC2, PIK3R1, YWHAZ, PIK3R5, HSP90B1, RELN, COL4A2, PIK3CA, KIT, SGK3, COL6A5, COL9A2, EIF4E2, TP53*	22	3.4	9.90E-03
MAPK signaling pathway	*MAP2K3, NTRK1, MAP2K4, CSF1, CACNA2D2, BRAF, CACNA1D, CACNA1E, CACNA1G, DUSP7, CACNA1I, IL1B, KIT, NF1, RAPGEF2, TP53, MYD88,*	19	3	1.60E-02

	MAP4K4, MAP3K5			
Apoptosis	*NTRK1, DFFA, BCL2A1, HTRA2, PIK3R1, PTPN13, ERN1, PIK3CA, AIFM1, ATM, TP53, BIRC2, MAP3K5*	13	2	2.40E-03
Rap1 signaling pathway	*MAP2K3, CSF1, FPR1, BRAF, PIK3R1, RAP1GAP, PLCB4, PIK3CA, RASSF5, KIT, KRIT1, CTNNB1, RAPGEF2*	13	2	5.70E-02
Phospholipase D signaling pathway	*GRM5, PLCB4, PIK3CA, KIT, GAB1, TSC2, PIP5K1A, PIK3R1, PIK3R5, AGPAT4*	10	1.6	6.80E-02
Cellular senescence	*PIK3CA, CHEK2, RASSF5, TSC2, ATM, CACNA1D, E2F4, PIK3R1, TP53*	10	1.6	8.70E-02
ECM-receptor interaction	*RELN, COL4A2, LAMA2, VWF, LAMB2, SV2A, COL9A2, COL6A5, FREM2*	9	1.4	1.10E-02
Protein digestion and absorption	*COL15A1, SLC36A2, CTRL, COL4A2, COL11A1, COL20A1, COL9A2, COL6A5, SLC8A1*	9	1.4	2.50E-02
Growth hormone synthesis, secretion and action	*MAP2K3, MAP2K4, STAT5B, PLCB4, PIK3CA, IRS4, CACNA1D, PIK3R1, SSTR5*	9	1.4	5.30E-02
Acute myeloid leukemia	*BCL2A1, PIK3CA, ZBTB16, KIT, BRAF, PIK3R1, RUNX1T1*	8	1.2	7.70E-03
Central carbon metabolism in	*NTRK1, PIK3CA, KIT, PIK3R1, SLC16A3, TP53,*	8	1.2	9.80E-03

cancer	*GCK, HK2*			
Prostate cancer	*HSP90AA1, PIK3CA, CTNNB1, BRAF, PIK3R1, TP53, ETV5, HSP90B1*	8	1.2	4.80E-02
Inflammatory mediator regulation of TRP channels	*NTRK1, MAP2K3, PLCB4, PIK3CA, IL1B, PRKCQ, PIK3R1, ASIC3*	8	1.2	5.10E-02
Toll-like receptor signaling pathway	*MAP2K3, IFNAR2, MAP2K4, PIK3CA, IL1B, CD80, PIK3R1, MYD88*	8	1.2	6.60E-02
TNF signaling pathway	*MAP2K3, MAP2K4, PIK3CA, CSF1, IL1B, PIK3R1, BIRC2, MAP3K5*	8	1.2	9.60E-02
GnRH secretion	*GABBR2, CACNA1I, PLCB4, PIK3CA, CACNA1D, PIK3R1, CACNA1G*	7	1.1	2.20E-02
Cortisol synthesis and secretion	*CACNA1I, MC2R, PLCB4, CYP11B1, CACNA1D, PBX1, CACNA1G*	7	1.1	2.30E-02
ErbB signaling pathway	*MAP2K4, STAT5B, PIK3CA, GAB1, BRAF, PIK3R1, PAK5*	7	1.1	7.10E-02
Small cell lung cancer	*COL4A2, LAMA2, PIK3CA, LAMB2, PIK3R1, TP53, BIRC2*	7	1.1	9.50E-02
Starch and sucrose metabolism	*AMY2B, AGL, ENPP1, GYG2, GCK, HK2*	6	0.9	7.20E-03
Carbohydrate digestion and absorption	*PLCB4, PIK3CA, AMY2B, CACNA1D, PIK3R1, HK2*	6	0.9	2.20E-02

Drug-gene interaction

A comprehensive analysis of drug-gene interactions has identified a number of genes with potential as druggable targets (Table 24; Figure 15). These genes include *ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16,* MT-*CO3* and *TLL1*. Each of these genes underscores their suitability as targets for future drug development and therapeutic interventions.

Table 24.: Potential Druggable genes identified in somatic genes in TNBC.

Gene	Interaction_types	Drug_name
THBD		SIMVASTATIN
THBD		CILOSTAZOL
THBD		LEVOTHYROXINE
THBD		GINKGO
THBD		CURCUMIN
THBD		ALPROSTADIL
RYR1	activator, channel blocker	RYANODINE
RYR1	activator, antagonist	cA2
RYR1	channel blocker	PROCAINE
RYR1	channel blocker	RUTHENIUM RED
RYR1	antagonist	MAGNESIUM

RYR1	antagonist	DANTROLENE
RYR1	activator	SURAMIN
RYR1	activator	ADENOSINE TRIPHOSPHATE
RYR1		DANTROLENE
RYR1	antagonist	DANTROLENE SODIUM
RYR1		DANTROLENE
RYR1	activator	CAFFEINE
THBD		SIMVASTATIN
THBD		CILOSTAZOL
THBD		LEVOTHYROXINE
THBD		GINKGO
THBD		CURCUMIN
THBD		ALPROSTADIL
TP53		epirubicin
TP53		Trametinib
TP53	vaccine	AD.P53-DC
TP53		Cisplatin
TP53		YONDELIS
TP53	activator	DCL000015

TP53		Pramlintide
TP53		CAPECITABINE
TP53		Docetaxel
TP53		PROPYLTHIOURACIL
TP53		PUROMYCIN
TP53		Crizotinib
TP53		RAPAMYCIN
TP53		AZD6738
TP53		p53-SLP vaccine
TP53		Irinotecan

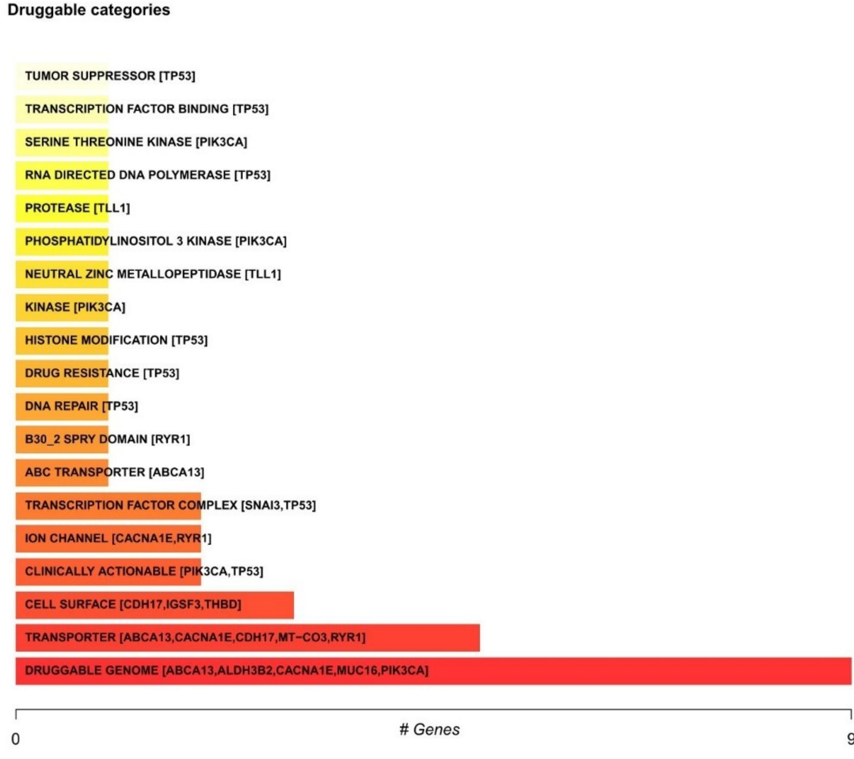


Figure 15. Druggable genes and druggable categories identified in somatic TNBC.

Germline mutation identification

In a germline variant analysis, 557 variants across the genome were identified, spanning 152 different genes after applying variant filtration to focus on the most clinically relevant ones (Table 25). The variant filtration criteria were targeting exonic, non-synonymous and pathogenic variants only, resulting in a subset of 557 missense mutations. Furthermore, single-nucleotide polymorphism (SNP) mutations were the most frequently observed, followed by deletions (Figure 16 A and B).

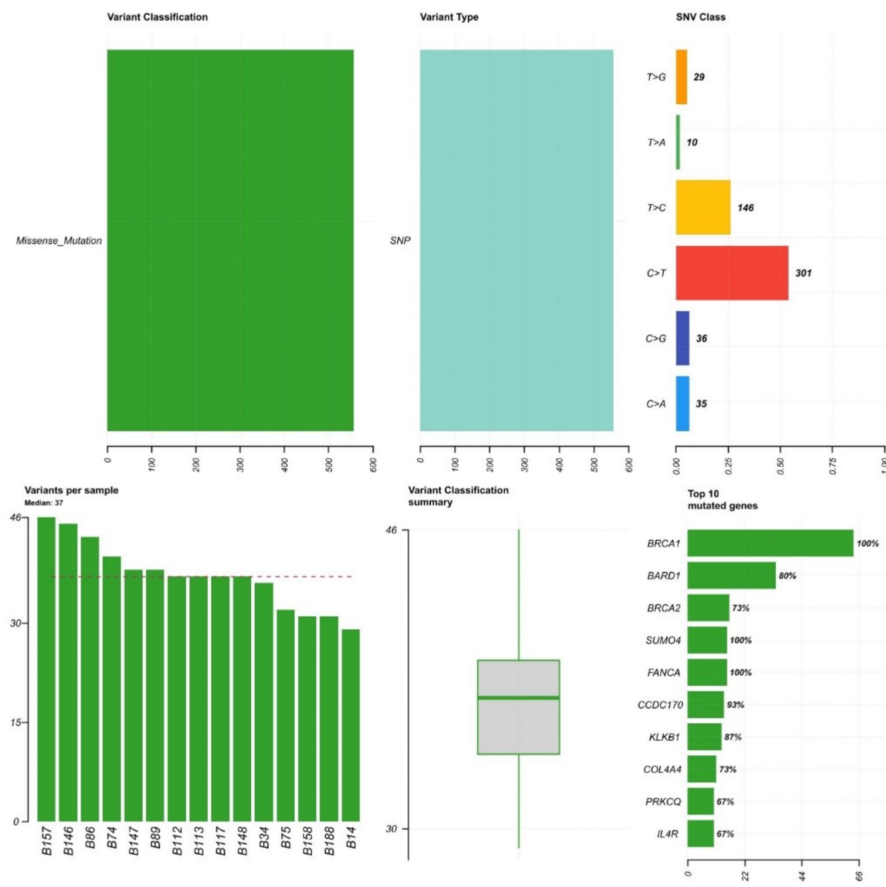


Figure 16. Mutational landscape in germline variants in TNBC.

Notably, C > T mutations, categorized as pathogenic or of unknown significance, constituted the predominant class among single-nucleotide variants (Figure 16 C). The number or variants per sample ranged from 29 to 46, with a median value of 37 (Figure 16 D). Among the 10 most mutated genes, Maftools flags MUC16 as a gene that is usually found mutated in exome studies, therefore likely to be a passenger gene.

Table 25. Summary of germline variants analysis.

ID	summary
Samples	15
Number of Genes	152
Missense Mutation	557
Total	557

Frequently mutated genes in germline variants

The germline whole exome sequencing analysis of TNBC analysis identified a large number of genetic variants found in the germline blood samples, showing the predominant genes frequently altered. The most frequently mutated genes were *BRCA1, SUMO4, FANCA, BARD1, BRCA2, CCDC170, KLKB1, PRKCQ, COL4A4,* and *IL4R* (Table 26). Among these genes, the prevalence of genetic variations varied from patient to patient. Notably, *BRCA1, SUMO4*, and *FANCA* displayed the highest prevalence of common genetic variants and short indels present in all patients (100%) from the tumor samples. Following this, *CCDC170* displayed a prevalence of 93% in the samples, while *KLKB1, BARD1, BRCA2, COL4A4, PRKCQ,* and *IL4R* displayed frequencies of 87%, 80%, 73%, 73%, 67%, and 67%, respectively (Figure 16 D).

Table 26: Top 10 frequently mutated genes and the number of Missense mutations and altered samples in germline TNBC.

Hugo Symbol	Missense Mutation	Altered Samples
BRCA1	64	15
FANCA	15	15
SUMO4	15	15
CCDC170	14	14
KLKB1	13	13
BARD1	34	12
BRCA2	16	11
COL4A4	11	11
IL4R	10	10
PRKCQ	10	10

Interacting gene pair analysis

A comprehensive analysis was conducted, focusing on the relationship between gene pairs within germline TNBC (Triple-Negative Breast Cancer) patients to identify whether the genes commonly affected in these patients tend to appear together or if their occurrences are mutually exclusive (Table 27). Interestingly, two gene pairs, *PRKCQ-BARD1* and *APOB-NQO1*, were identified to display a significant tendency to co-occur, with calculated p values of 0.05 (Figure 17). This suggests a potential interdependence or mutual association between these pairs, signifying a correlated role or interaction in the context of TNBC patients. Co-occurring genes could imply that mutations in one gene increase the likelihood or impact of mutations in another, pointing toward synergistic effects or shared pathways contributing to the condition. Additionally, six gene pairs, such as *F5-ARSA, MTHFR-GHRL, F5-BRIP1, TTN-GPSM2, APOB-MTHFR* and *GPSM2-IL4R* were highlighted by our findings to exhibit a contrasting pattern by being mutually exclusive with calculated p values of 0.05. This indicates a tendency for these genes to avoid occurring together within the same samples of TNBC patients.

Table 27: Significantly co-occuring gene pairs and mutually exclusive gene pairs in germline TNBC.

Gene1	Gene2	Event	p-Value	p-Adj
PRKCQ	*BARD1*	Co_Occurence	0.022	0.425
APOB	*NQO1*	Co_Occurence	0.040	0.561
F5	*ARSA*	Mutually_Exclusive	0.007	0.154
MTHFR	*GHRL*	Mutually_Exclusive	0.028	0.498
F5	*BRIP1*	Mutually_Exclusive	0.040	0.561
TTN	*GPSM2*	Mutually_Exclusive	0.040	0.561
APOB	*MTHFR*	Mutually_Exclusive	0.040	0.561
GPSM2	*IL4R*	Mutually_Exclusive	0.044	0.575

The observations, as illustrated in Figure 3, shed light on the intricate relationships between specific gene pairs within the germline TNBC context. Identifying mutually exclusive genes can offer insights into different pathways underlying a particular condition or trait. For instance, mutations in different genes in cancer may be mutually exclusive, which indicates that they perform similar functions within one specific pathway while mutations in one gene may prevent mutations in another. Understanding the patterns of co-occurrence and mutual exclusivity among these genes is crucial for unravelling the underlying mechanisms driving TNBC and could potentially inform future therapeutic strategies or diagnostic approaches.

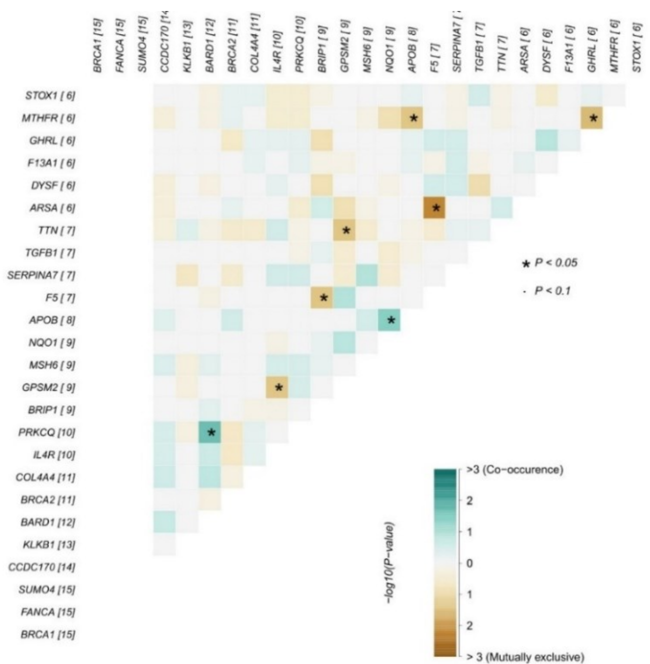


Figure 17. Interacting gene pair analysis of the 30 most mutated genes in germline TNBC.

Mutational signature analysis in germline

Mutational signature analysis in germline TNBC revealed the presence of three distinct mutational signatures associated with 75 genes (Figure 18; Table 28). Accordingly, the signature 1 consisted of 26 genes with a cosine similarity score of

0.608 demonstrate remarkable congruence with the COSMIC_6 mutational signature with an aetiology of defective DNA mismatch repair. Similarly, the signature 2 included other 26 genes show a cosine similarity score of 0.721 with the mutational signature of COSMIC_6 also of defective DNA mismatch repair aetiology. Furthermore, the signature 3 consisted of another subset of three genes match the COSMIC_1 mutational signature, scoring 0.713 on the cosine similarity scale. A defective DNA mismatch repair mechanism was found to be responsible for the etiology of these genes.

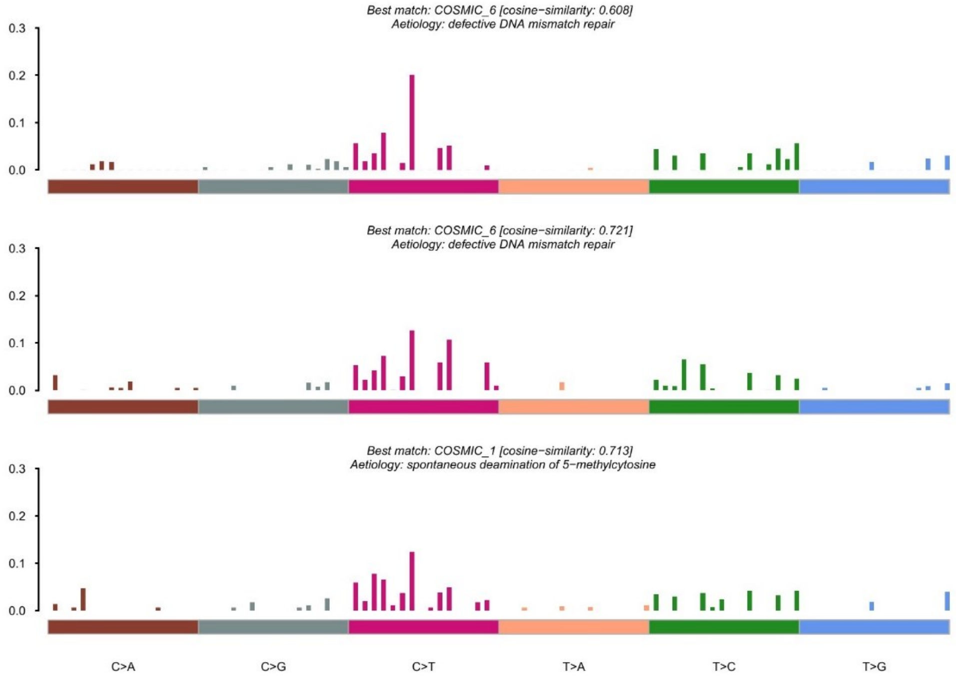


Figure 18. Mutational signature in germline TNBC matching with COSMIC signature database.

Table 28. Mutational signature genes match with COSMIC database in germline TNBC.

Hugo Symbol	Group1	Group2	Number of	Number of	P-value

			mutated group1	**mutated group2**	
KLKB1	Signature_1	Rest	3 of 5	10 of 10	0.095
PRKCQ	Signature_1	Rest	5 of 5	5 of 10	0.101
SERPINA7	Signature_1	Rest	4 of 5	3 of 10	0.118
BRIP1	Signature_1	Rest	2 of 5	7 of 10	0.328
ARSA	Signature_1	Rest	3 of 5	3 of 10	0.328
F13A1	Signature_1	Rest	3 of 5	3 of 10	0.328
GHRL	Signature_1	Rest	3 of 5	3 of 10	0.328
BARD1	Signature_1	Rest	5 of 5	7 of 10	0.505
BRCA2	Signature_1	Rest	3 of 5	8 of 10	0.561
MTHFR	Signature_1	Rest	1 of 5	5 of 10	0.581
ZNF469	Signature_1	Rest	1 of 5	5 of 10	0.581
IL4R	Signature_1	Rest	4 of 5	6 of 10	0.601
APOB	Signature_1	Rest	2 of 5	6 of 10	0.608
TTN	Signature_1	Rest	3 of 5	4 of 10	0.608
COL4A4	Signature_1	Rest	4 of 5	7 of 10	1
DYSF	Signature_1	Rest	2 of 5	4 of 10	1
STOX1	Signature_1	Rest	2 of 5	4 of 10	1
BRCA1	Signature_1	Rest	5 of 5	10 of 10	1
FANCA	Signature_1	Rest	5 of 5	10 of 10	1
SUMO4	Signature_1	Rest	5 of 5	10 of 10	1
CCDC170	Signature_1	Rest	5 of 5	9 of 10	1
GPSM2	Signature_1	Rest	3 of 5	6 of 10	1
MSH6	Signature_1	Rest	3 of 5	6 of 10	1
NQO1	Signature_1	Rest	3 of 5	6 of 10	1
F5	Signature_1	Rest	2 of 5	5 of 10	1
TGFB1	Signature_1	Rest	2 of 5	5 of 10	1

MTHFR	Signature_2	Rest	5 of 6	1 of 9	0.011
GHRL	Signature_2	Rest	0 of 6	6 of 9	0.027
PRKCQ	Signature_2	Rest	2 of 6	8 of 9	0.089
SERPINA7	Signature_2	Rest	1 of 6	6 of 9	0.118881
MSH6	Signature_2	Rest	2 of 6	7 of 9	0.135664
ZNF469	Signature_2	Rest	4 of 6	2 of 9	0.135664
DYSF	Signature_2	Rest	1 of 6	5 of 9	0.286713
BRIP1	Signature_2	Rest	5 of 6	4 of 9	0.286713
APOB	Signature_2	Rest	2 of 6	6 of 9	0.314685
TGFB1	Signature_2	Rest	4 of 6	3 of 9	0.314685
IL4R	Signature_2	Rest	3 of 6	7 of 9	0.328671
CCDC170	Signature_2	Rest	5 of 6	9 of 9	0.4
KLKB1	Signature_2	Rest	6 of 6	7 of 9	0.485714
BARD1	Signature_2	Rest	4 of 6	8 of 9	0.525275
BRCA2	Signature_2	Rest	5 of 6	6 of 9	0.604396
F5	Signature_2	Rest	2 of 6	5 of 9	0.608392
GPSM2	Signature_2	Rest	3 of 6	6 of 9	0.622378
NQO1	Signature_2	Rest	3 of 6	6 of 9	0.622378
ARSA	Signature_2	Rest	3 of 6	3 of 9	0.622378
BRCA1	Signature_2	Rest	6 of 6	9 of 9	1
FANCA	Signature_2	Rest	6 of 6	9 of 9	1
SUMO4	Signature_2	Rest	6 of 6	9 of 9	1
COL4A4	Signature_2	Rest	4 of 6	7 of 9	1
TTN	Signature_2	Rest	3 of 6	4 of 9	1
F13A1	Signature_2	Rest	2 of 6	4 of 9	1
STOX1	Signature_2	Rest	2 of 6	4 of 9	1
APOB	Signature_3	Rest	4 of 4	4 of 11	0.076923
MSH6	Signature_3	Rest	4 of 4	5 of 11	0.103297

ARSA	Signature_3	Rest	0 of 4	6 of 11	0.103297
MTHFR	Signature_3	Rest	0 of 4	6 of 11	0.103297
DYSF	Signature_3	Rest	3 of 4	3 of 11	0.235165
GHRL	Signature_3	Rest	3 of 4	3 of 11	0.235165
F5	Signature_3	Rest	3 of 4	4 of 11	0.282051
TTN	Signature_3	Rest	1 of 4	6 of 11	0.569231
TGFB1	Signature_3	Rest	1 of 4	6 of 11	0.569231
GPSM2	Signature_3	Rest	3 of 4	6 of 11	0.604396
NQO1	Signature_3	Rest	3 of 4	6 of 11	0.604396
F13A1	Signature_3	Rest	1 of 4	5 of 11	0.604396
ZNF469	Signature_3	Rest	1 of 4	5 of 11	0.604396
BRIP1	Signature_3	Rest	2 of 4	7 of 11	1
STOX1	Signature_3	Rest	2 of 4	4 of 11	1
BRCA1	Signature_3	Rest	4 of 4	11 of 11	1
FANCA	Signature_3	Rest	4 of 4	11 of 11	1
SUMO4	Signature_3	Rest	4 of 4	11 of 11	1
CCDC170	Signature_3	Rest	4 of 4	10 of 11	1
KLKB1	Signature_3	Rest	4 of 4	9 of 11	1
BARD1	Signature_3	Rest	3 of 4	9 of 11	1
BRCA2	Signature_3	Rest	3 of 4	8 of 11	1
COL4A4	Signature_3	Rest	3 of 4	8 of 11	1
IL4R	Signature_3	Rest	3 of 4	7 of 11	1
PRKCQ	Signature_3	Rest	3 of 4	7 of 11	1
SERPINA7	Signature_3	Rest	2 of 4	5 of 11	1

Pathway Analysis

A pathway analysis revealed a variety of oncogeneic pathways associated with the genes including Pathways in cancer, PI3K-Akt signaling pathway, Human papilloma virus infection, Platinum drug resistance, Hypertrophic cardiomyopathy, Gastric cancer, Breast cancer, Colorectal cancer, FoxO signaling pathway, Mismatch repair, Central carbon metabolism, PD-L1 expression and PD-1 check point pathway in cancer and thyroid cancer (Table 29).

Table 29: Oncogenic pathways associated with variants and their related genes in germline TNBC.

KEGG_PATHWAY	Genes	Count	%	Pvalue
Pathways in cancer	*NTRK1,RET,NQO1,WNT10A,TGFB1,IL4R,NOTCH1,JUP,IFNGR1,LAMA1,PTCH1,LRP5,BRCA2,MLH1,EGFR,MSH6,CDH1,MSH3,COL4A4,COL4A5*	20	13.15	1.47E-05
PI3K-Akt signalling pathway	*NTRK1,RET,IL4R,LAMA1,ITGB3,TSC1,BRCA1,EGFR,STK11,RELN,COL4A4,COL4A5,COL6A3*	13	8.55	0.001173
Human papilloma virus infection	*WNT10A,RELN,NOTCH1,LAMA1,COL4A4,ITGB3,COL4A5,COL6A3,TSC1,ATM,TCIRG1,EGFR104*	12	7.89	0.001983
Platinum drug resistance	*MSH6,ATP7B,MSH3,ATM,BRCA1,MLH1,POLH*	7	4.60	2.29E-04

Gastric cancer	*WNT10A,TGFB1,JUP,CDH1, LRP5,MLH1,EGFR*	7	4.60	0.008888
Breast cancer	*WNT10A,NOTCH1,LRP5, BRCA1,BRCA2,EGFR*	6	3.94	0.031582
Colorectal cancer	*MSH6,TGFB1,MSH3,MLH1, EGFR*	5	3.28	0.019591
FoxO signalling pathway	*STK11,TGFB1,PRKAG2, ATM,EGFR*	5	3.28	0.072293
Mismatch repair	*MSH6,MSH3,POLD1,MLH1*	4	2.63	0.002494
Central carbon metabolism in cancer	*NTRK1,RET,PGAM2,EGFR*	4	2.63	0.051512
PD-L1 expression and PD-1 check point pathway in cancer	*CD4,IFNGR1,PRKCQ, EGFR*	4	2.63	0.091087
Thyroid cancer	*NTRK1, RET, CDH1*	3	1.97	0.072371

Drug-gene interaction

Drug–gene interactions analysis in germline samples identified eighteen genes as potential druggable targets, including *BRCA1, NQO1, BRIP1, IL4R, COL4A4, FANCA, TTN, F5, APOB, TGFB1, KLKB1, BRCA2, ARSA, SERPINA7, BARD1, PRKCQ, GPSM2* and *MSH6* (Figure 19). These genes are potential targets for future drug development and therapeutic interventions. A rich assortment of genes offers opportunities for further research and possible breakthroughs in precision medicine, promising tailored treatments and a better understanding of their role in physiological and pathological processes.

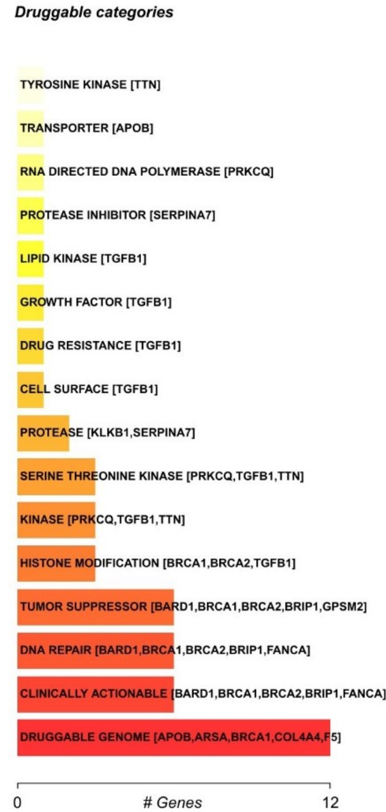


Figure 19. Druggable genes and druggable categories identified in germline variants in TNBC.

Clinical Data Analysis

In the clinical data analysis, a comprehensive examination revealed numerous somatic variants and genes strongly correlated with distinct clinical features. This investigation involved a comprehensive comparison between contrasting clinical traits, examining those that were either present or absent for a specific characteristic. By analyzing these differing traits, a focused effort was made to identify the specific genes and variants associated with these distinctive clinical attributes. The analysis

concentrated on discerning the genetic elements linked to the contrasting characteristics observed in the clinical data. By specifically comparing the presence or absence of these traits, the study aimed to isolate the genes and variants that aligned with each contrasting clinical feature. This approach allowed for a more targeted and precise identification of the genetic components associated with each distinct clinical trait.

The analysis highlighted the prevalence of specific genes and variants related to clinical features, with 39 variants associated with Age at menarche, 36 variants linked to Co morbidities, and another 36 variants concerning the distinction between the Left or Right Breast. Additionally, 34 variants were correlated with salt intake, while 33 variants were found associated with Location, and 30 variants with Grade, signifying their prominence among the factors examined. However, within this cohort study, the analysis did not indicate any significant correlations between Alcohol intake, Age at menopause, and First-degree family history of breast cancer, suggesting a lack of substantial association within this specific dataset. This extensive scrutiny of clinical attributes and their genetic correlations presents a comprehensive understanding of the interconnectedness between genetic variants and distinct clinical features, shedding light on the pronounced influence of specific genetic elements on various aspects of breast cancer-related attributes within the studied cohort.

Table 30: Summary of clinical data analysis displaying clinical characteristics and corresponding variant counts in identified genes.

Factors	**Number of genes**
Age at menarche	39
Co morbidities	36
Left Right Breast	36
Salt intake	34
Location	33
Grade	30
DCIS	26
Mutton Beef Pork	24

Age	23
Exercise	22
Saum	18
Kuhva/ Gutkha/ Sahdah/ Khaini/ Tuibur/ Smoking	18
Recurrence	16
Age at first delivery	15
Number of children	12
Breast feeding	11
Familial or Non Familial	10
Impression	9
No of Lymph nodes	9
Water intake day	4
Second degree family history of breast cancer	4
Survival months	4
Alive or Deceased	3
Fish Chicken	Single factor
Fruits Vegetable	Single factor
Smoked meat and vegetables	Single factor
Any first or second degree family history of ovarian cancer?	Single factor
Lymph nodes	Single factor
Alcohol	No significant associations
Age at menopause	No significant associations
First degree family history breast cancer	No significant associations
Size of grossed tumour in cm	No significant associations

CHAPTER 6

DISCUSSION

Breast cancer is a complex disease, meaning that a number of variables can contribute to its development. While the fact that the disease is global, there are significant racial and geographic differences in its incidence, treatment response, mortality, and survival rates. These differences may be caused by different kinds of variables, like environment, genetics, lifestyle, and population structure. The prevalence of breast cancer has increased due to changes in risk factors, which the number is increasing everyday. By classifying women according to their risk factors for breast cancer, risk-free practices can be improved, and targeted programs for breast cancer screening may exist (Momenimovahed and Salehiniya, 2019).

Pathological and genetic investigations are the gold standard for accurate diagnosis of cancers. This study will try to identify several genes that associate to TNBC and its association with risks and overall breast cancer for improved cancer risk assessment. The information on gene-specific risks for TNBC is useful to identify the genes. This study is important as it aims to characterize the immunological profile and mutations. It will also help to find out the pervasiveness of mutations unique to Mizo ethnic groups and TNBC. This may help in better clinical management of individuals at risk for or diagnosed with the cancer.

In Mizo population, about 67.96% of the cases had breast cancer when they were older than 40. Exercise, sleeping patterns, or night work do not appear to be significantly correlated with tumor grade according these studies. Tumor grade may be influenced by the frequency or amount of consumption of any particular variable, including pig, fish, beef, chicken, fruits, vegetables, Sa-um (fermented pork fat), smoked meat, smoked vegetables, oil, salt, and water.

An analysis of the relationship between the consumption of alcohol and tobacco and tumor grade revealed no significant results for any of the variables (Zodinpuii et al., 2020).

With 65 cases out of 129 control samples, the age range of 41 to 51 years has the highest risk factors. Obese women may have a higher than average chance of developing breast cancer (Babiker et al., 2020). A significant p-value of 0.000 was found for BMI > 30, indicating that 42.3% of the cases were obese. Increased breast cancer cases are significantly correlated with higher BMI. In contrast, family history is quite significant. 5.2% of respondents who were control subjects have a history of breast cancer, and only 5.6% of them were married (Babiker et al., 2020).

About 80% of cases of breast cancer, including TNBCs, occur in people older than 50. Risk increases to 1.5% at age 40, 3% at **age** 50, and over 4% at age 70 years (Almansour, 2022). Women's breast cells are especially susceptible to hormonal abnormalities including the **hormones** progesterone and estrogen. The circulation of androgens and estrogen has been associated to a higher risk of breast cancer. The profile of a BRCA2-associated tumor is correlated with Luminal B subtype, whereas that of BRCA1-related tumors is similar to the TNBC subtype. TNBC is still very common in non-Hispanic white women. Additionally, black women are thought to have the lowest rates of cancer survival, and their death rate is substantially greater. Like TNBC, a major risk factor for breast cancer is one's **family history**. A first-degree breast cancer relative is reported by 13–19% of patients with breast cancer diagnoses (Almansour, 2022).

Compared to postmenopausal and white women, African American and premenopausal women had a significantly greater chance of developing basal-like BC and a significantly reduced chance of developing luminal A BC. The study population had a higher TNBC prevalence than white BC patients. TNBC was independently correlated with younger age, premenopausal status, increased parity, use of hormonal contraceptives, high histological grade, and advanced disease. The body mass index (BMI) and HR expression were inversely correlated in premenopausal women. On the other hand, BMI positively correlated with HR and HER2 levels in postmenopausal women (Jiao et al., 2014).

This study investigated into the comprehensive analysis of whole exome sequencing data sourced from 15 patients diagnosed with triple-negative breast cancer in Mizoram. The dataset encompassed 15 tissue cancer samples, corresponding adjacent normal tissues, and 15 blood (germline) samples. We initiated our study by performing Whole Exome Sequencing (WES) analysis on somatic tissue samples of Triple-Negative Breast Cancer (TNBC). Subsequently, we investigated into germline variants analysis. These comprehensive analyses encompassed Single Nucleotide Variant analysis, identification of frequently mutated genes, exploration of mutually exclusive gene patterns, deciphering mutational signatures, assessing drug-gene interactions, and analysing pathways involved. Furthermore, our investigation extended to clinical data analysis pertaining to somatic TNBC samples, leading to the discovery of single nucleotide variants (SNVs) and their related genes associated with TNBC.

WES analysis of the somatic TNBC samples identified a total of 188 somatic single nucleotide variants (SNVs) and revealed associations with 647 genes. These variants exhibited diverse types, with the most prevalent being nonsense mutations (560), followed by frame shift mutations (37) and several other mutation categories. The germline WES analysis in blood samples unveiled 557 variants within 152 genes after filtering for exon, non-synonymous mutations and pathogenic variants only. Notably, among the identified variants, amino acid substitutions in proteins, largely attributed to genetic disorders, were predominantly linked to missense mutations arising from single nucleotide variations in both cases. This observation highlights the significance of these diverse mutations in influencing genetic anomalies, with missense mutations, in particular, holding a substantial role in protein-level alterations and consequent biological implications. The comprehensive breakdown of these findings unveils a significant understanding of the genetic landscape within the context of potential genetic disorders and offers a basis for further investigation into their functional and clinical implications (Loewe & Hill, 2010).

This analysis of somatic variants unveiled TP53 as the predominant gene exhibiting a notably high frequency of genetic variations within the tumor samples studied. Notably, 47% of the patients displayed mutations in TP53 that signifies the prevalence of TP53 mutations in somatic breast cancer patients, whereas the remaining 15 genes collectively showcased a 13% mutation frequency, signifying their relevance in the landscape of genetic alterations observed in these samples. The diverse array of genes, including *CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD*, and *TLL1,* collectively contribute to the mutational landscape. On the other hand, the most frequently mutated genes in germline samples are *BRCA1, SUMO4*, and *FANCA* that displayed the highest prevalence of common genetic variants and short indels present in all patients (100%). Following this, *CCDC170* displayed a prevalence of 93% in the samples, while *KLKB1, BARD1, BRCA2, COL4A4, PRKCQ,* and *IL4R* displayed frequencies of 87%, 80%, 73%, 73%, 67%, and 67%, respectively. These findings are consistent with previous studies, underscoring the consistency and relevance of these genetic variations in breast cancer research (Ahmad et al., 2023; Kaur et al., 2023b; Luo et al., 2021)

Additionally, significant co-occurrences were identified between two pairs of genes in somatic TNBC samples *OTULINL/CDH17* and *MUC16/MT-CO3*, signifying associations with p-values of 0.05 whereas, none of the genes are mutually exclusive. On the other hand, in germline samples, a significant co-occurrence tendency was observed between *PRKCQ-BARD1* and *APOB-NQO1*. The presence of these pairs suggests a possible interaction or interdependence between these two pairs indicating that they may be interdependent or mutually associated in TNBC. Mutations in one gene may increase the probability or impact of mutations in another, suggesting that there may be synergistic effects or shared pathways. Additionally, six gene pairs, such as *F5-ARSA, MTHFR-GHRL, F5-BRIP1, TTN-GPSM2, APOB-MTHFR* and *GPSM2-IL4R*, appeared to exhibit a contrasting pattern by being mutually exclusive, suggesting that these genes tend not to occur together in the same germline TNBC samples. For instance, in cancer, mutations in different genes may be mutually exclusive, which

implies that their functions are similar within one pathway, whereas mutations in one gene might prevent mutations in another gene.

Furthermore, the mutational signature analysis in the somatic TNBC samples identified three distinctive signatures associated with TP53 gene. Signature 1 and 2 match COSMIC_6 mutational signature with an aetiology of defective DNA mismatch repair. DNA mismatch repair (MMR) is responsible for preventing mutations and genetic instability by identifying and repairing mismatches or small distortions that may occur during DNA replication. The signature 3 closely match the COSMIC_3 mutational signature with an aetiology of defects in DNA−DSB repair by HR (pathway to repair DNA double-strand breaks). These DNA repair problems could contribute to genetic instability and cancer-related diseases (Aparicio et al., 2014). On the other hand, the mutational signature analysis in the germline TNBC samples revealed the presence of three distinct mutational signatures associated with 75 genes. Signature 1 consisted of 26 genes and Signature 2 consisted of another 26 genes, both of them congruent with the COSMIC_6 mutational signature with a defective DNA mismatch repair aetiology. In addition, signature 3 consisted of another subset of three genes, which matched the COSMIC_1 mutational signature and had an etiology of a defective DNA mismatch repair mechanism. DNA mismatch repair (MMR) is responsible for preventing mutations and genetic instability by identifying and repairing mismatches or small distortions that may occur during DNA replication.

Pathway analysis in somatic samples showed several genes associated with the oncogenic pathways including PI3K-Akt signalling pathway, MAPK signalling pathway, Apoptosis, Rap1 signalling pathway, Phospholipase D signalling pathway, and Cellular senescence. Whereas, in the oncogenic pathways identified in germline samples include Pathways in cancer, PI3K-Akt signalling pathway, Human papilloma virus infection, Platinum drug resistance, Hypertrophic cardiomyopathy, Gastric cancer, Breast cancer, Colorectal cancer, FoxO signalling pathway, Mismatch repair, Central

carbon metabolism in cancer, PD-L1 expression and PD-1 check point pathway in cancer and thyroid cancer.

Drug-gene interactions have been extensively investigated, and thirteen genes have been identified as potentially druggable targets for somatic samples, including *ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16, MT-CO3* and *TLL1*. In germline samples, eighteen genes have been identified as possible druggable targets, including *BRCA1, NQO1, BRIP1, IL4R, COL4A4, FANCA, TTN, F5, APOB, TGFB1, KLKB1, BRCA2, ARSA, SERPINA7, BARD1, PRKCQ, GPSM2* and *MSH6.* These genes are frequently mutated genes and may potentially targets for future drug development and therapeutic interventions. (Al-Shamsi et al., 2016; Derynck et al., 1985; Ellis et al., 2012; Herman et al., 2012; Joenje & Patel, 2001; Kogan & Carpizo, 2016; Miki et al., 1994; Wooster et al., 1995). A rich assortment of genes offers opportunities for further research and possible breakthroughs in precision medicine, promising tailored treatments and a better understanding of their role in physiological and pathological processes

The analysis of clinical features along with the somatic TNBC samples identified several variants and their associated genes including 39 variants associated with Age at menarche, 36 variants linked to co-morbidities, and another 36 variants concerning the distinction between the left or right breast cancer. Additionally, 34 variants were correlated with salt intake, while 33 variants were found associated with Location, and 30 variants with Grade, signifying their prominence among the factors examined. Nevertheless, within this cohort study, the analysis did not indicate any significant correlations between Alcohol intake, Age at menopause, and First-degree family history of breast cancer, suggesting a lack of substantial association within this specific dataset. In contrast, since there are only 15 samples, some clinical characteristics may not be sufficient to make accurate predictions due to the limited sample size. However, the study of clinical attributes and their genomic variants provide a clear understanding of how genetic variants are interconnected with distinct clinical features, revealing the

significant impact of certain genetic elements on various aspects of breast cancer-related characteristics.

Cancer remains a global health challenge, necessitating continuous research to unravel its intricacies and develop effective therapeutic strategies. One fundamental aspect of cancer etiology is the accumulation of genetic mutations, both in somatic and germline cells, which contribute to the initiation and progression of malignancies. In this study, we provide a comprehensive account of the results obtained from an in-depth analysis of somatic mutations, germline variants, mutational signatures, and their impact on crucial oncogenic pathways. We explore how genetic diversity contributes to the complexity of this disease and integrate clinical data to identify genes or genetic variations linked to clinical outcomes. We also integrate clinical data to identify genes or variants related to the particular effect of the clinical causes. We defined the genetic alterations in 15 triple negative breast cancers using a whole exome sequencing data with a clinical data analysis including somatic and germline data from Mizoram, India. The most frequent somatic alterations were mutations in *TP53* (47%) and *ELAPOR1* (13%) and *CACNA1E* (13%) and other genes at 13% mutation rates among the top 10 mutated genes. Germline mutation revealed that among top mutated genes are *BRCA1* (100%), *SUMO4* (100%), *FANCA1* (100%), *CCDC170* (93%), *BARD1* (80%) *BRCA2* (73%). The mutational signature analysis revealed the presence of three distinct mutational signatures associated with 20 genes including the etiology of a defective DNA mismatch repair mechanism in COSMIC database. Furthermore, our pathway analysis revealed that it involves somatic variants include many oncogenic pathways are altered including PI3K-Akt signaling pathway, MAPK signaling pathway, Apoptosis and Rap1 signaling pathway these are the main altering pathway for carcinogenesis. This study represent the first whole exome exome sequence data of somatic and germline mutation analysis from Mizoram triple negative breast cancer patient.

CHAPTER 7

SUMMARY

- Somatic cancer analysis identified a total of 688 variants spanning 647 genes and the most frequently mutated genes are *TP53, CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD,* and *TLL1*.
- *TP53* exhibited the highest mutation rate, exceeding 43%, while the other genes had mutations around 7%.
- These findings highlight the significant role of *TP53* mutations in breast cancer among these patients.
- Germline variant analysis identified 557 variants in 152 genes after filtering for exon and non-synonymous mutations.
- *BRCA1, FANCA*, and *SUMO4* mutations were found in all blood samples.
- Risk factors for TNBC include young age at breast cancer diagnosis, young age at menarche, young age at time of first child birth, high parity, lack or shorter duration of breast feeding, premenopausal women with high body mass index and African American ethnicity and an elevated waist: hip ratio (Dawood, 2010).
- Some of the risk factors for TNBC are BRCA mutation, ethnicity, age and Body Mass Index (BMI) (Mousavi et al., 2019).
- TNBC has a worse prognosis, high recurrence and poor survival rates when compared to other subtypes of breast cancer (Dawood, 2010).
- Some of the risk factors for TNBC are BRCA mutation, ethnicity, age and Body Mass Index (BMI) (Mousavi et al., 2019).

- TNBC has a worse prognosis, high recurrence and poor survival rates when compared to other subtypes of breast cancer (Dawood, 2010).
- PALB2 has been identified as one of the common predisposing genes for breast cancer afer BRCA1/2 with penetrance estimated at 33–70% depending on age at diagnosis and family history (Siraj et al. 2023).
- BRCA1 and BRCA2 germline mutations are responsible for about 60% of HBCs with an overall 60–80% lifetime risk; the other 40% are associated with other predisposing variants in moderate-to-high penetrance genes such as PALB2, PTEN, TP53, CDH1, CHEK2, ATM, and the MMR group (Catana et al., 2023).
- 32.8% patients carried a pathogenic or likely pathogenic heterozygous germline mutation in 19 genes. including high-penetrant breast cancer genes like BRCA1 (43%), BRCA2 (19%), PALB2 (12%), TP53 (3%), and CDH1 (2%) (Catana et al., 2023).
- Eight genes were significantly mutated in the Breast cancer tumors including 54 PIK3CA, TP53, GATA3, MAP3K1, CDH1, CBFB, PTEN, and RUNX1 (Ding et al., 2023)
- EGFR pathway contains well-established oncogenes including EGFR, KRAS, BRAF and PIK3CA genes that modulate gene activations in solid tumors including lung, colorectal cancer (CRC) and pancreatic ductal adenocarcinoma (Zakaria et al. 2019).
- In our study, Significant gene co-occurrences were observed, including *OTULINL/CDH17* and *MUC16/MT-CO3,* with p-values of 0.05.
- Mutational signature analysis revealed three distinct signatures associated with 20 genes, with SBS 6 indicating defective DNA mismatch repair.
- Drug-gene interaction analysis identified potential druggable genes, including

ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16, MT-CO3, and *TLL1*.

- Clinical data analysis identified key factors contributing to the cohort's outcomes, such as age at menarche, comorbidities, breast location, salt intake, and grade. However, no significant associations were found for alcohol intake, age at menopause, or first-degree family history of breast cancer.

Appendix 1. Questionnaire for Studies on Gene mutations and Epidemiology among Breast cancer patient in the Mizo population

PERSONAL INFORMATION

Name (Hming): ____________________Male/Female (Mipa/Hmeichhia):_________

Age (Kum):______ Date of Birth (Pian Kum) :_________ Ph no:__________ Marital status (Nupui nei/ pasal nei/ neilo) ___________ Age at marriage: ________

Spouse Name(Nupui/Pasal hming) :____

No. of siblings (Pianpui unau):____________ Male (Mipa)__________ Female(Hmeichhia) ____________

Blood group:__________ Height (san zawng):____________ Weight (Rihzawng):__________

Birth place (Pianna khua):____

Migration history:____

Present Address (tuna awmna):____

HABITS

How often do you take exercise? (Exercise I la ngai em?)
Rarely/never ( ) Once a week ( ) 3-4 times a week ( ) Everyday ( )

How many hours of sleep do you get? (Ni khatah darkar engzat nge I mut?)
1- 3 hrs ( ) 3-6 hrs ( ) 6-8 hrs ( ) 8-10 hrs ( )

Is your job stressful or do you perform shift work (night duty)? (I hnathawh a hahthlak em?Zan lamah te hna I thawk thin em?) YES ( ) NO ( )

Meat consumption per week (Kar khata sa ei tam lam):

	Ei / Ei lo	Once	Twice	Thrice	Everyday
Pork(vawksa)					

Fish(sangha)					
Chicken(Arsa)					
Mutton(Kelsa)					
Others					

How often do you eat each of the following food? (A hnuaia chaw te hi engtiangin nge I ei that?)

Fruit /Fruit juices (Thei/Thei tui) :Never / Rarely / Occasionally / Normal / Regularly

Vegetables (thlai) : Never / Rarely / Occasionally / Normal / Regularly

Saum : Never / Rarely / Occasionally / Normal / Regularly

Smoked meat (Sa rep) : Never / Rarely / Occasionally / Normal / Regularly

Smoked Vegetables (Thlai rep): Never / Rarely / Occasionally / Normal / Regularly

Salt used (Chi hman):

Salt intake (Chi ei tam lam) : Never / Less / Normal / Heavy

Name of cooking oil used(Tel hman hming):

Amount of oil intake(Tel ei tam lam): Never / Less / Normal / Heavy

Water intake per day (ni khata tui in zat): 1-2 glass/500ml-1 ltr/1-2 ltrs / >2 ltrs

TOBACCO AND ALCOHOL HISTORY

Bettle nut (Kuhva Hring): Never / Less / Normal / Heavy

Gutkha (Zarda/supari/etc): Never / Less / Normal / Heavy

Sahdah: Never / Less / Normal / Heavy

Khaini: Never / Less / Normal / Heavy

Tuibur: Never / Less / Normal / Heavy

Do you smoke? If yes, which brand and how often? (Zozial / Beedi / Cigarette)
(Mei I zu em?I zu anih chuan eng anga tam nge I zuk thin?) Zuk tam zawng ______

Do you consume alcohol? If yes, which brand and how often?(Local/Branded /Both)
(Zu I in thin em?I in thin cuan eng anga zing/tam nge?): Never/Occasionally/
Normal/Regularly

REPRODUCTIVE HISTORY:
Age at Menarche (Thi neih tan kum):
No. of children (Fa neih zat):
Age at first delivery (Fa hmasa ber a pian a i kum zat):
Breast feeding (Hnute i pe em?):
Duration of Breast feeding (Hnute pek rei hunchhung):
Birth control pills (Indanna):
Abortions (Nau I ti tla tawh em?):
Age at menopause (Thi hul kum):

MEDICAL HISTORY

Have you ever been diagnosed with any other type of cancer? If yes,what type of cancer?(Cancer hrim hrim I vei tawh em? I vei tawh chuan eng cancer nge?)

Do you have any major illnesses?If yes,what type of illness?(Natna dang I nei em?I neih chuan eng natna nge?)

Have you done X-ray or CT scan? If yes, why? YES ( ) NO ( )
(X-ray emaw CT scan I ti tawh em?I tih tawh chuan engnge a chhan?)

FAMILY DETAILS (in relation to cancer)

Do you have any first-degree relatives - mother, sisters, daughters – with breast cancer?
(I nu, laizawn, fanu emaw hnute cancer vei an awm em?)

Do you have any second degree relatives diagnosed with breast cancer?
(I laina/cousin emaw hnute cancer vei an awm em?)

Do you have any 1st or 2nd degree relatives diagnosed with ovarian cancer?
(I chhungte chhul cancer vei an awm tawh em?)

Do you have any first or second degree relatives diagnosed with any other types of cancer?
(I chhungte dang cancer vei an awm tawh em,eng cancer nge?)

Any other type of major inheritable diseases in the family ?
(I chhungte dang natna hlauhawm inthlahchhawn theih vei an awm em?)

When was your cancer diagnosed?(Engtikah nge cancer I vei tih I hmuhchhuah?)

____

How do you suspect youself of Breast Cancer? (Engtinge I in rinhlelh?) ____

On which side of the breast was tumor found? Right ( ) Left ( ) Both sides ( ) (I hnute khawi lamah nge bawk an hmuhchhuah?)

After you were diagnosed with breast cancer, what type of treatment did you take? (I hnute cancer hmuhchhuah anih atang khan eng enkawlna nge I dawn?)

(Surgery / Chemotherapy / Radiation / Hormone Therapy / Any other)

Do you have a history of Fibroadenoma before you were diagnosed as breast cancer?( )Cancer i nih I in hriatchhuah hma in, I hnute emaw hnute bulah bawk a awm em?

HORMONE RECEPTOR STATUS

ER - (+) (-) PR - (+) (-) HER2 - (+) (-)

CONSENT (Remtihna)

The information provided above was given with my full consent and I do not have any objection in providing my biological sample for research purposes. I have read and understand the consent information.

(Heng a chunga thu te hi ka hriatpui a, ka biological sample hi zir chian atan pek ka remti thlap e.)

Place (Hmun): Signature:

Date(Ni): Name (Hming):

Ka lawm e

(THANK YOU VERY MUCH FOR YOUR HELP)

Appendix 2. Ethical Clearance forms

IEC, CIVIL HOSPITAL, AIZAWL.
COMMUNICATION TO THE PRINCIPAL INVESTIGATOR/GUIDE BY THE MEMBER SECRETARY, INSTITUTIONAL ETHICS COMMITTEE

No.B.12018/1/13-CH(A)/IEC/33 Dated: 15th of October, 2014

To,

Dr. R. Muthukumaran,
Associate Prof., Department of Chemistry,
Mizoram University.

The Institutional Ethics Committee in its meeting held on 14th of October 2014, has reviewed and discussed your application to conduct the clinical trial/project entitled; **Structural and Functional Characterization of Clinically Significant Mutation in h – BRCA-1 Gene Among the Mizoram Population.**

Sponsored by : **Department of Biotechnology, New Delhi.**

Code no. : ______________________

The following documents were reviewed:

a. Trial Protocol (including protocol amendments)/project, dated 4th of August, 2014.
b. Investigator's Brochure, dated 4th of August, 2014.
c. Patient Information Sheet and Informed Consent Form (including updates if any) in Hindi, English and/or vernacular language.
d. Proposed methods for patient accrual including advertisement (s) etc. proposed to be used for the purpose.
e. Current CV of investigator from outside Civil Hospital, Aizawl.
f. Insurance Policy/Compensation for participation and for serious adverse events occurring during the study participation.
g. Investigator's Agreement with the Sponsor,
h. Investigator's Undertaking.
i. Ethics Committee Proforma.
j. DCGI approval letter/submission letter, if any.
k. Civil Hospital, Aizawl Case Report Form
l. Any other/additional documents

Decision of Committee: **Approved/~~NotAprroved~~**

(DR. C. LALCHHANDAMA)
Member Secretary
Institutional Ethics Committee

INSTITUTIONAL ETHICS COMMITTEE
MIZORAM STATE CANCER INSTITUTE
ZEMABAWK. AIZAWL
(REGD NO. EC/NEW/INST/2020/1166, 19.11.2020)

No.D.12016/2/2013-MSCI/IEC/ Dated Aizawl, the 16th April 2021

APPROVAL NO: 2/2021

To

Prof N Senthil Kumar
Department of Biotechnology
Mizoram University

Dear Sir,

The Institutional Ethics Committee, Mizoram State Cancer Institute in its meeting held on date 16th April 2021, time 2:00 PM at Conference Hall, Mizoram State Cancer Institute reviewed your application for the study

"WHOLE GENOME ANALYSIS AND ITS PUTATIVE CLINICAL IMPACT ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION (PHARMACOGENES IN RELATION TO ADVERSE DRUG REACTION"

The Institutional Ethics Committee is glad to inform you that your application has been approved in **its present form and as per your declaration** given in the application to undertake / continue the study at Mizoram State Cancer Institute, Zemabawk, Aizawl, Mizoram.

Your Sincerely

16/4/21
(Dr C.LALCHHANDAMA)

Chairman
Institutional Ethics Committee
Mizoram State Cancer Institute
Zemabawk, Aizawl

16/4/21
(Dr B.ZOTHANKIMA)

Member Secretary
Institutional Ethics Committee
Mizoram State Cancer Institute
Zemabawk, Aizawl

1000 Genomes Project Consortium; Auton, A., Brooks, L. D., Durbin, R. M. et al (2015). A global reference for human genetic variation. Nature, 526 (7571): 68-74.

Abdollahi, A. and Etemadi, M. (2016) Pathological characteristics of triple-negative breast cancer at main referral teaching hospital, April 2014 to April 2015, Tehran, Iran. International Journal of Hematology-Oncology and Stem Cell Research, 10(4): 200–205.

Agrawal, N., Frederick, M.J., Pickering, C.R. et al (2011) Exome sequencing of head and necksquamous cell carcinoma reveals inactivating mutations in NOTCH1. Science, 333(6046): 1154–7.

Ahmad, H., Ali, A., Ali, R. et al (2023). Mutational Landscape and In-Silico Analysis of TP53, PIK3CA, and PTEN in Patients with Breast Cancer from Khyber Pakhtunkhwa. ACS Omega, 8(45): 43318-43331.

Alghamdi, I. G., Hussain, I., Alghamdi, M. et al (2015). Early marriage is a potential risk factor for female breast cancer in the Eastern Region of Saudi Arabia. American Journal of Research Communication, 3(7): 23.

Almansour N. M. (2022). Triple-Negative Breast Cancer: A Brief Review About Epidemiology, Risk Factors, Signaling Pathways, Treatment and Role of Artificial Intelligence. Frontiers in molecular biosciences, 9: 836417.

Al-Shamsi, H. O., Jones, J., Fahmawi, Y. et al (2016). Molecular spectrum of KRAS, NRAS, BRAF, PIK3CA, TP53, and APC somatic gene mutations in Arab patients with colorectal cancer: Determination of frequency and distribution pattern. Journal of Gastrointestinal Oncology, 7(6): 882–902.

Anderson, W. F. and Matsuno, R. (2006). Breast cancer heterogeneity: A mixture of at least two main types? In Journal of the National Cancer Institute, 98(14): 948–951.

Andrews, S. (2010). FastQC: a quality control tool for high throughput sequence data.

Angius, A., Pira, G., Cossu-Rocca, P. et al (2023). Deciphering clinical significance of BCL11A isoforms and protein expression roles in triple-negative breast cancer subtype. Journal of Cancer Research and Clinical Oncology, 149(7): 3951-3963.

Antony, M., Surakutty, B., Vasu, T. et al (2018). Risk factors for breast cancer among Indian women: A case-control study. Nigerian Journal of Clinical Practice, 21(4): 436–442.

Aparicio, T., Baer, R., and Gautier, J. (2014). DNA double-strand break repair pathway choice and cancer. DNA repair, 19: 169–175.

Babiker, S., Nasir, O., Alotaibi, S. H. et al (2020). Prospective breast cancer risk factors prediction in Saudi women. Saudi journal of Biological Sciences, 27(6): 1624–1631.

Benjamin, D., Sato, T., Cibulskis, K. et al (2019). Calling Somatic SNVs and Indels with Mutect2. bioRxiv 861054

Bhutia, Y. D., Ogura, J., Grippo, P. J. et al (2020). Chronic exposure to excess iron promotes EMT and cancer via p53 loss in pancreatic cancer. Asian journal of pharmaceutical sciences, 15(2): 237–251.

Brewer, H. R., Jones, M. E., Schoemaker, M. J. at al (2017). Family history and risk of breast cancer: an analysis accounting for family structure. Breast cancer research and treatment, 165: 193-200.

Burstein, M. D., Tsimelzon, A., Poage, G. M. et al (2015) Comprehensive Genomic Analysis Identifies Novel Subtypes and Targets of Triple-negative Breast Cancer Analysis and interpretation of data: HHS Public Access. Clinical cancer research : an official journal of the American Association for Cancer Research, 21(7): 1688–1698.

Cătană, A., Trifa, A. P., Achimas-Cadariu, P. A. et al (2023). Hereditary Breast Cancer in Romania—Molecular Particularities and Genetic Counseling Challenges in an Eastern European Country. Biomedicines, 11(5): 1386.

Chen, J. Q. and Russo, J. (2009). ERalpha-negative and triple negative breast cancer: molecular features and potential therapeutic approaches. Biochimica et biophysica acta, 1796(2): 162–175.

Chen, S., Zhou, Y., Chen, Y. et al (2018). Fastp: An ultra-fast all-in-one FASTQ preprocessor. Bioinformatics, 34(17): i884–i890.

Chivukula, M., Striebel, J. M., Ersahin, Ç. et al (2008). Evaluation of morphologic features to identify "basal-like phenotype" on core needle biopsies of breast. Applied Immunohistochemistry & Molecular Morphology, 16(5): 411-416.

Dawood, S. (2010) Triple-negative breast cancer: Epidemiology and management options. Drugs, 70(17): 2247–2258.

De Ruijter, T. C., Veeck, J., De Hoon, J. P. J. et al (2011) Characteristics of triple-negative breast cancer. Journal of Cancer Research and Clinical Oncology, 137(2): 183–192.

Derynck, R., Jarrett, J. A., Chen, E. Y. et al (1985). Human transforming growth factor-β complementary DNA sequence and expression in normal and transformed cells. Nature, 316(6030): 701–705.

Díaz-Velásquez, C. E., Gitler, R., Antoniano, A. et al (2023). Evaluation of genetic alterations in hereditary cancer susceptibility genes in the Ashkenazi Jewish women community of Mexico. Frontiers in genetics, 14: 1094260.

Ding, Y., Ding, K., He, X. et al (2023). The Value of Neoadjuvant Anthracycline-Based Regimens for HER2-Positive Breast Cancer: A Systematic Review and Meta-analysis Including 1366 Patients. Clinical Medicine Insights: Oncology, 17: 11795549231195293.

Dorling L, Carvalho S, Allen J. et al (2021) Breast Cancer Risk Genes - Association Analysis in More than 113,000 Women. New England Journal of Medicine, 384(5): 428-439.

Ellis, M. J., Ding, L., Shen, D., Luo, J. et al (2012). Whole-genome analysis informs breast cancer response to aromatase inhibition. Nature, 486(7403): 353–360.

Forbes, S. A., Bindal, N., Bamford, S. et al (2011) COSMIC: Mining complete cancer genomes in the catalogue of somatic mutations in cancer. Nucleic Acids Research, 39: 945–950.

Ghatak, S., Lallawmzuali, D., Lalmawia. et al (2014) Mitochondrial D-Loop and Cytochrome OxidaseC Subunit I Polymorphisms Among The Patients Breast Cancer Of Mizoram, Northeast India. Current Genetics, 60(3): 201– 212.

Gibbs, R. A., Boerwinkle, E., Doddapaneni, H. et al (2015) A global reference for human genetic variation. Nature, 526(7571): 68–74.

Griffith, M., Spies, N. C., Krysiak, K. et al (2017) CIViC is a community knowledgebase for expert crowdsourcing the clinical interpretation of variants in cancer. Nature Genetics, 49(2): 170–174.

Hanahan, D. and Weinberg, R. A. (2011). Hallmarks of cancer: the next generation. Cell, 144(5): 646–674.

Herman, D. S., Lam, L., Taylor, M. R. G. et al (2012). Truncations of Titin Causing Dilated Cardiomyopathy. New England Journal of Medicine, 366(7): 619–628.

Hu, C., Hart, S. N., Gnanaolivu, R. et al (2021). A population-based study of genes previously implicated in breast cancer. New England Journal of Medicine, 384(5): 440-451.

Irvin, W. J. and Carey, L. A. (2008). What is triple-negative breast cancer? European Journal of Cancer, 44(18): 2799–2805.

Jian, W., Shao, K., Qin, Q. et al (2017). Clinical and genetic characterization of hereditary breast cancer in a Chinese population. Hereditary cancer in clinical practice, 15: 19.

Jiao, Q., Wu, A., Shao, G. et al (2014). The latest progress in research on triple negative breast cancer (TNBC): risk factors, possible therapeutic targets and prognostic markers. Journal of thoracic disease, 6(9): 1329–1335.

Joenje, H., and Patel, K. J. (2001). The emerging genetic and molecular basis of Fanconi anaemia. In Nature Reviews Genetics, 2(6): 446–457.

Jones, M. E., Schoemaker, M. J., Wright, L. B. et al (2017). Smoking and risk of breast cancer in the Generations Study cohort. Breast Cancer Research, 19: 1-14.

Kaur, J., Chandrashekar, D. S., Varga, Z. et al (2023). Whole-Exome Sequencing Reveals High Mutational Concordance between Primary and Matched Recurrent Triple-Negative Breast Cancers. Genes, 14(9): 1690.

Kogan, S. and Carpizo, D. (2016). Pharmacological targeting of mutant p53. Translational cancer research, 5(6): 698–706.

Kotsopoulos, J., Moody, J. R., Fan, I. et al (2012). Height, weight, BMI and ovarian cancer survival. Gynecologic oncology, 127(1): 83-87.

Kulothungan, V., Sathishkumar, K., Leburu, S. et al (2022). Burden of cancers in India - estimates of cancer crude incidence, YLLs, YLDs and DALYs for 2021 and 2025 based on National Cancer Registry Program. BMC cancer, 22(1): 527.

Landrum, M. J., Lee, J. M., Riley, G. R. et al (2014) ClinVar: Public archive of relationships among sequence variation and human phenotype. Nucleic Acids Research, 42(D1): 980–985.

Landry, I., Sumbly, V., and Vest, M. (2022). Advancements in the Treatment of Triple-Negative Breast Cancer: A Narrative Review of the Literature. Cureus, 14(2): e21970.

Lee, A., Mavaddat, N., Wilcox, A. N. et al (2019). BOADICEA: a comprehensive breast cancer risk prediction model incorporating genetic and nongenetic risk factors. Genetics in medicine: official journal of the American College of Medical Genetics, 21(8): 1708–1718.

Li, H. and Durbin, R. (2009). Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics, 25(14): 1754–1760.

Li, H., Handsaker, B., Wysoker, A. et al (2009). The Sequence Alignment/Map format and SAMtools. Bioinformatics, 25(16): 2078–2079.

Li, Y. W., Xu, J., Zhu, G. Y. et al (2018). Apigenin suppresses the stem cell-like properties of triple-negative breast cancer cells by inhibiting YAP/TAZ activity. Cell Death Discovery, 4(1): 105.

Liu, S., Lachapelle, J., Leung, S., Gao, D., Foulkes, W. D., & Nielsen, T. O. (2012). CD8+ lymphocyte indafiltration is an independent favorable prognostic indicator in basal-like breast cancer. Breast cancer research *: BCR*, *14*(2): R48.

Lodha, R. S., Nandeshwar, S., Pal, D. K. et al (2011). Risk factors for breast cancer among women in Bhopal urban agglomerate: a case-control study. Asian Pac J Cancer Prev, 12(8): 2111-5.

Loeb, K.R. and Loeb, L.A. (2000). Significance of multiple mutations in cancer. Carcinogenesis, 21 (3): 379–385.

Loewe, L. and Hill, W. G. (2010). The population genetics of mutations: good, bad and indifferent. Philosophical transactions of the Royal Society of London. Series B, Biological sciences, 365(1544): 1153–1167.

Łukasiewicz, S., Czeczelewski, M., Forma, A. et al (2021). Breast Cancer-Epidemiology, Risk Factors, Classification, Prognostic Markers, and Current Treatment Strategies-An Updated Review. Cancers, 13(17): 4287.

Luo, R., Chong, W., Wei, Q. et al (2021). Whole-exome sequencing identifies somatic mutations and intratumor heterogeneity in inflammatory breast cancer. NPJ breast cancer, 7(1): 72.

Malvia, S., Bagadi, S. A., Dubey, U. S. et al (2017). Epidemiology of breast cancer in Indian women. Asia-Pacific Journal of Clinical Oncology, 13(4): 289-295.

Mathur, P., Sathishkumar, K., Chaturvedi, M. et al (2020). Cancer Statistics, 2020: Report From National Cancer Registry Programme, India. JCO global oncology, 6: 1063–1075.

Mayakonda, A., Lin, D. C., Assenov, Y. et al (2018). Maftools: Efficient and comprehensive analysis of somatic variants in cancer. Genome Research, 28(11): 1747–1756.

McKenna, A., Hanna, M., Banks, E. et al (2010). The genome analysis toolkit: A MapReduce framework for analyzing next-generation DNA sequencing data. Genome Research, 20(9): 1297–1303.

Mehrotra, R., and Yadav, K. (2022). Breast cancer in India: Present scenario and the challenges ahead. World Journal of Clinical Oncology, 13(3): 209–218.

Miki, Y., Swensen, J., Shattuck-Eidens, D. et al (1994). A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science (New York, N.Y.), 266(5182): 66–71.

Moinfar F. (2007) Essentials of Diagnostic Breast Pathology. A Practical Approach Springer, 2007th edition: pp496

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

Mousavi, S. A., Kasaeian, A., Pourkasmaee, M. et al (2019) Assessing the prognostic factors, survival, and recurrence incidence of triple negative breast cancer patients, a single center study in Iran. PloS one, 14(1), e0208701.

Nofech-Mozes, S., Trudeau, M., Kahn, H. K. et al (2009) Patterns of recurrence in the basal and non-basal subtypes of triple-negative breast cancers. Breast Cancer Research and Treatment, 118(1): 131–137.

Pachuau, L., Zami, Z., Nunga, T. et al (2022). First-degree family history of cancer can be a potential risk factor among head and neck cancer patients in an isolated Mizo tribal population, northeast India. Clinical Epidemiology and Global Health, 13: 100954.

Palachandra, A., Ishwaraprasad, G. D., Sreelatha, C. Y. et al (2017). Risk factors associated with carcinoma breast: a case control study. International Surgery Journal, 4(9): 3136-3140.

Pieńkowski, T. and Zielinski, C. C. (2009) Trastuzumab treatment in patients with breast cancer and metastatic CNS disease. Annals of Oncology, 21(5): 917–924.

Pierobon, M. and Frankenfeld, C. L. (2013) Obesity as a risk factor for triple-negative breast cancers: A systematic review and meta-analysis. Breast Cancer Research and Treatment, 137(1): 307–314.

Rhee, J., Han, S. W., Oh, D. Y. et al (2008) The clinicopathologic characteristics and prognostic significance of triple-negativity in node-negative breast cancer. BMC

cancer, 8: 307.

Sha, S., Si, L., Wu, X. et al (2022). Prognostic analysis of cuproptosis-related gene in triple-negative breast cancer. Frontiers in immunology, 13: 922780.

Shanker, N., Mathur, P., Das, P. et al (2021). Cancer scenario in North-East India & need for an appropriate research agenda. The Indian journal of medical research, 154(1): 27–35.

Sherman, B. T., Hao, M., Qiu, J. et al (2022). DAVID: a web server for functional enrichment analysis and functional annotation of gene lists (2021 update). Nucleic Acids Research, 50(W1): W216–W221.

Shimelis, H., LaDuca, H., Hu, C. et al (2018) Triple-negative breast cancer risk genes identified by multigene hereditary cancer panel testing. Journal of the National Cancer Institute, 110(8): 855–862.

Sinha, S., Nath, J., Mukherjee, A. et al (2016). Predictive and prognostic factors in breast cancer and their association with ER PR HER2/neu expression. J Carcinog Mutagen, 7(2): 1-4.

Siraj, A. K., Bu, R., Parvathareddy, S. K. et al (2023). PALB2 germline mutations in a large cohort of Middle Eastern breast-ovarian cancer patients. Scientific Reports, 13(1): 7666.

Sirisena, N. D., Adeyemo, A., Kuruppu, A. I. et al (2018). Genetic variants associated with clinicopathological profiles in sporadic breast cancer in Sri Lankan women. Journal of breast cancer, 21(2), 165-172.

Sofi, G. N., Sofi, J. N., Nadeem, R. et al (2012) Estrogen Receptor and Progesterone Receptor Status in Breast Cancer in Relation to Age, Histological Grade, Size of Lesion and Lymph Node Involvement. Asian Pacific Journal of Cancer Prevention, 13(10): 5047–5052.

Somasundaram K. (2010). BRCA1 and BRCA1 Genes and Inherited Breast and/or Ovarian Cancer: Benefits of Genetic Testing. Indian journal of surgical oncology, 1(3): 245–249.

Sørlie, T., Perou, C. M., Tibshirani, R. et al (2001) Gene expression patterns of breast

carcinomas distinguish tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America, 98(19): 10869–10874.

Suba, Z. (2014). Triple-negative breast cancer risk in women is defined by the defect of estrogen signaling: preventive and therapeutic implications. OncoTargets and therapy, 147-164.

Tetreault, M., Bareke, E., Nadaf, J., Alirezaie, N., & Majewski, J. (2015). Whole-exome sequencing as a diagnostic tool: current challenges and future opportunities. Expert review of molecular diagnostics, 15(6): 749–760.

Wang, K., Li, M. and Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research, 38(16): e164.

Whitesell, M. (2014) Cultural and Environmental Factors of Triple-Negative Breast Carcinomas. Cancer Research Journal, 2(2): 18.

Wilkinson, L., and Gathani, T. (2022). Understanding breast cancer as a global health concern. The British journal of radiology, 95(1130): 20211033.

Wood, E. H. (2004) The National Comprehensive Cancer Network (NCCN). Journal of the Medical Library Association, 92(3): 382–383.

Wooster, R., Bignell, G., Lancaster, J. et al (1995). Identification of the breast cancer susceptibility gene BRCA2. Nature, 378(6559): 789–792.

Wu, Q., Siddharth, S. and Sharma, D. (2021). Triple Negative Breast Cancer: A Mountain Yet to Be Scaled Despite the Triumphs. Cancers, 13(15): 3697.

Yadav, R.P., Ghatak, S., Chakraborty, P. et al (2018). Lifestyle chemical carcinogens associated with mutations in cell cycle regulatory genes increases the susceptibility to gastric cancer risk. Environmental science and pollution research international, 25(31): 31691–31704.

Yao-Lung, K., Dar-Ren, C. and Tsai-Wang, C. (2011). Clinicopathological features of triple-negative breast cancer in Taiwanese women. International journal of clinical oncology, 16: 500-505.

Yin, L., Duan, J. J., Bian, X. W. et al (2020). Triple-negative breast cancer molecular subtyping and treatment progress. Breast cancer research: BCR, 22(1): 61.

Zakaria, Z., Zulkifle, M. F., Wan Hasan, W. A. N. et al (2019). Epidermal growth factor receptor (EGFR) gene alteration and protein overexpression in Malaysian triple-negative breast cancer (TNBC) cohort. OncoTargets and therapy, 7749-7756.

Zodinpuii, D., Pautu, J. L., Zothankima, B. et al (2020). Clinical features and first-degree relative breast cancer, their correlation with histological tumor grade: a 5-year retrospective case study of breast cancer in Mizoram, India. Environmental science and pollution research, 27(2): 1991–2000.

Zomawia, E., Zami, Z., Vanlallawma, A. et al (2023). Cancer awareness, diagnosis and treatment needs in Mizoram, India: evidence from 18 years trends (2003-2020). The Lancet regional health. Southeast Asia, 17: 100281.

BIO-DATA

---

1. Name : Dr Lalawmpuii Pachuau
2. Full correspondence address: House No B-1/A;
YMA Road,
Chawnpui Veng, Aizawl,
Mizoram - 796009
3.Phone number : 9862379444
4. Email : maawmi1@yahoo.com
5. Date of birth : 15 May 1980
5. Gender : Female
6. Academic Qualification (Undergraduate Onwards)

	Degree	Institution	University	Year
1.	MBBS	Moti Lal Nehru Medical College	Allahabad	1998-2004
2.	MD Pathology	Regional Insitute of Medical Sciences	Manipur	2009-2012
3.	Senior Residency	Post Graduate Institute of Medical Education and Research, Chandigarh	Punjab	2012-2015

7. Work experience (in chronological order).

After obtaining post-graduation degree, I completed 3 years Senior Residency in PGIMER Chandigarh where we had rotator postings in Histopathology Department which includes daily reporting and making diagnosis of 60-80 cases, frozen section, immunohistochemistry and medical autopsy. Cytology and Gynaecology Department posting includes performing FNAC, radiological guided FNAC, body fluids and cervical PAP smear reporting. Hematology Department posting includes daily routine hematology reporting, performing Bone marrow aspiration and biopsy, staining IHC of bone marrow biopsy and reporting, flowcytometry and Cytogenetics posting.

Presently I am working as a Pathologist in the Department of Pathology, Civil Hospital Aizawl which is the main and busiest hospital in Mizoram where we get maximum number of cases from all over the State. Daily reporting of Hematology, Cytology, Histopathology and grossing is being done.

I enrolled as PhD scholar in the Department of Biotechnology, MZU in 2018 and my synopsis titled "Clinicopathological and Genomic Mutational Analysis for Triple Negative Breast Cancer in Mizo Population". I attended National and International conferences and CME regularly. Take active part in seminars, presenting papers and posters.

My personal interest is in Cancer Research and Genomics and get myself involved in cancer research. I was also sponsored by the Director of National Institute of Biomedical Genomics (NIBMG) Kalyani for Winter School on Circulating tumour cells.

List of Publications

Sailo CV, Zami Z, Ghatak S, Nemi L, Lalremmawia K, **Pachuau L**, Zomawia E, Siama Z, Kumar NS. Prevalence of High-Risk HPV Types in Women with Negative Cervical Cytology in a State of Northeast India with a High Burden of Cervical Cancer. **Indian Journal of Gynecologic Oncology**. 2022;20(1):8

Senthilkumar B, Zodinpuii D, **Pachuau L**, Chenkual S, Zohmingthanga J, Kumar NS, Hmingliana L. Ensemble Modelling for Early Breast Cancer Prediction from Diet and Lifestyle. **IFAC PapersOnLine**. 2022;55(1):429-35

Chakraborty P, Ghatak S, Chenkual S, **Pachuau L**, Zohmingthanga J, Bawihtlung Z, Khenglawt L, Pautu JL, Maitra A, Chhakchhuak L, Kumar NS. Panel of significant risk factors predicts early stage gastric cancer and indication of poor prognostic association with pathogens and microsatellite stability. **Genes and Environment.** 2021;43(1):1-5.

Lalchhandama C, **Pachuau L**, Lalchhanhimi T, Zohmingthanga J, Kumar NS. High incidence invasive breast cancer is associated with Her2/neu (+)/ER (+)/PR (+) and high grade ductal carcinoma in young age patients of Northeast India. **Current Medicine Research and Practice**. 2020;10(4):160-4.

Mukherjee S, Madathil SA, Ghatak S, Jahau L, Pautu JL, Zohmingthanga J, **Pachuau L**, Nicolau B, Kumar NS.. Association of tobacco smoke–infused water (tuibur) use by Mizo people and risk of Helicobacter pylori infection. **Environmental Science and Pollution Research.** 2020; 27:8580-5.

Zodinpuii D, Pautu JL, Zothankima B, **Pachuau L**, Kumar NS. Clinical features and first degree relative breast cancer, their correlation with histological tumor grade: a 5-year retrospective case study of breast cancer in Mizoram, India. **Environmental Science and Pollution Research**. 2020; 27:1991-2000.

Lalawmpuii P, Vanlalhlua C. Subcutaneous Cysticercosis mimicking soft tissue neoplasm – a case report. **IOSR- Journal of Dental and Medical Sciences (IOSR-JDMS)**. 2020;19(8):15-20.

Vanlalhlua C, **Lalawmpuii P**. Enterobius Vermicularis Appendicitis – A case report. **IOSR Journal of Dental and Medical Sciences (IOSR-JDMS).** 2020;19(2):11-14.

Eric Zomawia, Zothan Zami, Andrew Vanlallawma, Nachimuthu Senthil Kumar, John Zothanzama, Lalchhanhimi Tlau, Lalchhandama Chhakchhuak, **Lalawmpuii Pachuau**, Jeremy L. Pautu, and Evelyn V. L. Hmangaihzuali, Cancer awareness, diagnosis and treatment needs in Mizoram, India: evidence from 18 years trends (2003–2020). The **Lancet Regional Health Southeast Asia**. 2023; 17: 100281.

Krishnan Sathishkumar, Jayasankar Sankarapillai, Aleyamma Mathew, Rekha A. Nair, Nitin Gangane, Eric Zomawia, Tseten Wangyal Bhutia, Kaling Jerang, Preethi Sara George, Swapna Maliye, Rajesh Laishram, Anand Shah, Shiromani Debbarma, Shravani Koyande, **Lalawmpuii Pachuau** et al, Survival of patients with cervical cancer in India – findings from 11 population based cancer registries under National Cancer Registry Programme. **The Lancet Regional Health - Southeast Asia**. 2023; 100296.

Payel Chakraborty, Sillarine Kurkalang, Souvik Ghatak, Subrata Das, Arindam Palodhi, Sumanta Sarkar, Ranjan Dhar, Saia Chenkual, Lalawmpuii Pachuau, John Zohmingthanga, Jeremy L. Pautu, Thomas Zomuana, Sailo Tlau Lalruatfela, John Zothanzama, Nachimuthu Senthil Kumar, Arindam Maitra. Deep sequencing reveals recurrent somatic mutations and distinct molecular subgroups in Gastric Cancer in Mizo population, North East India. **Genomics** 2023 (In Press).

Conference Presentations

- **Lalawmpuii Pachuau**. *Immunohistochemical Profile of Invasive Breast Cancer in Mizo Female Patients*, 2nd Annual Convention of North East (India) Academy of Science and Technology (NEAST) & International Seminar on Recent Advances in Science and Technology (IRSRAST) during *16 – 18th November 2020* (Virtual) organized by NEAST, Mizoram University, Aizawl-796004, Mizoram (India).
- **Lalawmpuii Pachuau.** *Epidemiological Risk factors related to TNBC in Mizo Population.* 12th Annual Convention of Association of Pharmacy and Biotechnology (ABAP) & International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Trends (BEHIET 2018) at Mizoram University, Mizoram, India
- **Lalawmpuii Pachuau.** *Rheumatoid arthritis, Immunopathophysiology, and Diagnostics criteria***.** Mizoram orthopedic Society CME 2018.

Declaration:

I do hereby declare that the above-mentioned details are correct best of my knowledge. I will produce the necessary certificates whenever asked.

Date: 1 November 2023 Faithfully

Lalawmpuii Pachuau

PARTICULARS OF THE CANDIDATE

NAME OF THE CANDIDATE	:	Lalawmpuii Pachuau
DEGREE	:	Ph.D.
DEPARTMENT	:	Biotechnology
TITLE OF THE THESIS	:	Clinicopathological and Genomic Mutational Analysis of Triple Negative Breast Cancer in Mizo Population
DATE OF ADMISSION	:	07.08.2018
APPROVAL OF RESEARCH PROPOSAL		
1. DRC	:	03.05.2019
2. BOS	:	07.05.2019
3. SCHOOL BOARD	:	17.05.2019
MZU REGISTRATION NO.	:	1807391
Ph.D. REGISTRATION NO. & DATE	:	MZU/Ph.D./1292 of 07.08.2018
EXTENSION (IF ANY)	:	---

Head

Department of Biotechnology

ABSTRACT

CLINICOPATHOLOGICAL AND GENOMIC MUTATIONAL ANALYSIS OF TRIPLE NEGATIVE BREAST CANCER IN MIZO POPULATION

AN ABSTRACT SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENTS FOR THE DEGREE OF DOCTOR OF PHILOSOPHY

LALAWMPUII PACHUAU

MZU REGN NO.: 1807391

Ph.D. REGN NO.: MZU/Ph.D./1292 of 07.08.2018

DEPARTMENT OF BIOTECHNOLOGY

SCHOOL OF LIFE SCIENCES

NOVEMBER 2023

CLINICOPATHOLOGICAL AND GENOMIC MUTATIONAL ANALYSIS OF TRIPLE NEGATIVE BREAST CANCER IN MIZO POPULATION

BY

LALAWMPUII PACHUAU

Department of Biotechnology

Supervisor: Prof. N. SENTHIL KUMAR

Submitted

In partial fulfillment of the requirement of the degree of Doctor of Philosophy in Biotechnology of Mizoram University, Aizawl

INTRODUCTION

Cancer of the breast, lung and cervix are the top 3 cancers in Women in India in the year 2020. According to data collected by Population Based Cancer Research (PBCR) between 2010-2014, it is the third most common cancer among women in Mizoram. BC is the second most common cancer worldwide, comprising of about 10.4% of all cancers and is the second most common cause of cancer death in females. Breast cancer (BC) is a malignant tumor that starts in the breast cells and is typically seen in the lobules or ducts of the breast. Rarely, connective tissues can also develop into a site of breast cancer. When cancer cells leave the lymph node and spread to nearby and distant healthy breast tissues as well as other parts of the body, this is referred to as metastasis. The distance that cancer cells have spread from the tumor's origin determines the stage of breast cancer. The hormonal receptor expression in breast cancer (BC) is lower and the age at presentation is younger among Indians as compared to western population.

Breast cancer is a multifaceted and intricate illness that is influenced by a range of genetic, hormonal, and environmental variables, such as lifestyle choices, food habits, and reproductive history. Breast cancers are typically referred to as carcinomas, while they can also occasionally be called adenocarcinomas. Breast cancers can be classified into many categories based on the place and severity of the malignancy. "*In situ*" cancer means that it is still present at its original location; and "invasive or infiltrating" means that it has spread to neighboring tissues. The site of the cancer's onset was indicated by the names of the tissues or cells.

Breast cancer that is triple-negative or basal-like, lacks expression of the HER2/neu oncoprotein (HER2-) and the hormone receptor (ER-/PR-). BRCA1 gene mutations and younger women are commonly associated with triple negative breast cancer. Breast cancer is a heterogeneous disease characterized by dysregulation of multiple cellular pathways and have different sensitivities to treatment. Immunohistochemistry (IHC) for estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor 2 (HER2) expression is analyzed

on all breast tumour specimens which will decide the type of treatment that the patients will be receiving. Triple negative breast cancer (TNBC) is an aggressive subtype having distinct clinical and biological characteristic defined by absence of ER, PR, and HER2 expression.

Risk factors for TNBC include young age at breast cancer diagnosis, young age at menarche, young age at time of first child birth, high parity, lack or shorter duration of breast feeding, premenopausal women with high body mass index and African American ethnicity and an elevated waist:hip ratio. A small percentage of instances of breast cancer (5–10%) are inherited, and they are caused by mutations in the cancer susceptibility genes, BRCA1 and BRCA2, which are inherited autosomally. Just 10 - 20% of hereditary cases of breast cancer are caused by mutations in these two genes. There have also been reports of other genes, including TP53, PTEN, ATM, CHD1, CHEK2, SKT11, BRIP1, PALB2 and others having mutations that raise the risk of breast cancer. The hallmark of inherited breast cancer is often a younger age of cancer onset (~40 years).

About 90% of incidences of breast cancer are sporadic. A number of reproductive, environmental, and demographic variables, in addition to those modifier genetic variants with a minor increase in risk or those predicted low-to-moderate penetrance, are important in the development of breast cancer. The cancer usually manifests as unilateral and has a late age of initiation.

Mizoram, a state in the north-eastern region of India with a unique socio-cultural context, provides an excellent environment for studying breast cancer epidemiology. Furthermore, Mizoram had the highest incidence of cancer in India between 2003 and 2010, but this dropped to fourth in 2012-2014, according to The Mizoram Population-Based Cancer Registry (PBCR) which reported cancer cases in 2003 at the regional level was established for event monitoring.

The study aims to analyze the Clinicopathological and hormonal status of the Triple Negative Breast Cancer (TNBC) cases. Further, the genomic changes were screened through whole exome analysis. The findings of this study may also provide useful information on the prevalence of mutations and non-genetic factors as risk factors for the development of breast cancer in Mizo population.

OBJECTIVES

1) To analyze Triple Negative Breast Cancer (TNBC) with clinico-pathological parameters.
2) To evaluate the estrogen receptor (ER), progesterone receptor (PR) and HER-2/neu expression in invasive breast carcinomas by immunohistochemistry.
3) To perform whole exome analysis to characterize the landscape of genetic alterations underlying TNBC in Mizo population.

MATERIALS AND METHODS

Patients who had undergone modified radical mastectomy at Civil Hospital Aizawl were included in the study. The histological type, lymph node involvement, tumor grade, estrogen and progesterone receptors and HER-2/neu status were assessed for all the samples. Fresh as well as formalin fixed paraffin embedded (FFPE) breast tissue from the main tumour and adjacent normal tissue and corresponding blood samples was taken for analysis. A skilled technician extracted 2 ml of blood from the participants (patients and controls), the blood was placed in EDTA vials and stored at -20°C for further analysis.

Clinical records, age at diagnosis, gender, tumor type or grade, habits and familial history of the patients were collected through structured questionnaire. Samples were collected only with prior consent of the patients. A total of 240 samples of MRM specimen were collected and 59 patients are found to be Triple negative in immunohistochemistry and were included in the study. Whole Exome

Sequencing (WES) was done on 15 patients tumour, adjacent normal and whole blood sample.

The demography, reproductive history, environmental variables, and family history of the recruited individuals with regard to cancer and other hereditary disorders are among the epidemiological aspects taken into consideration. The patient's reproductive history was taken into consideration, including their age at marriage/ menarche/ menopause, parity, age at first delivery, marital status, no.of children, length of breastfeeding, usage of oral contraceptive pills, and history of abortion. Environmental factors include eating habits (fruits, vegetables, *sa-um*, smoked food, salt, water, and oil), sleeping schedules, night shift work, and exercise habits. Tobacco and alcohol intake are also factors (Betel nut, gutkha, sahdah, khaini, tuibur and cigarette).

Age-based frequency distribution of the study's cases and controls was computed. The relationship between demographic characteristics and TNBC risk was evaluated using chi-square testing. To calculate the possible confounder's impact of environmental variables on breast cancer, logistic regression analysis was performed. IBM Statistical Package for Social Sciences (SPSS), version 22.0 for Windows was used for statistical analyses. The odds ratio and confidence interval were calculated using MedCalc Statistical Software version 20.113 (https://www.medcalc.org/calc/odds_ratio.php).

Ethical approval for the study was obtained from Civil Hospital Aizawl (B.12018/1/13-CH(A)/IEC/33) and Mizoram State Cancer Institute- Human Ethical Committee.

Clinico-pathological parameters like patient's age, histological type, grade and lymph node metastasis

On microscopic examination, the histological type and tumour grade as per Modified Bloom Richardson score would be classified. In this scoring system, three factors that are taken into consideration and each of these factors are scored from 1-3.

Hormonal receptor expression by IHC

Modified radical mastectomy specimens from Surgery operation theatre, Civil Hospital Aizawl were grossed as per standard protocols and histopathological examination by haematoxylin-eosin staining for microscopic examination was done. Tumor characteristics regarding type of tumor and histological grade are classified as per modified Bloom Richardson histologic score which take into account the amount of gland formation, nuclear features or nuclear pleomorphism and mitotic activity of the tumor. ER, PR scoring will be done as per the Allred scoring system and HER-2/neu scoring will be done according to the standard reporting protocols.

Whole Exome Sequence analysis to characterize the landscape of genetic alterations underlying triple negative breast cancer (TNBC) in Mizo population.

Genomic DNA was isolated from frozen tumor tissues using AllPrep DNA Mini Kit (QIAGEN, Lot. 51304) according to manufacturer's instructions. Genomic DNA from Whole Blood was isolated using QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN) and stored in -20°C according to manufacturer's instructions. 20 µl of QIAGEN Protease was pipetted into 1.5 ml microcentrifuge tube.

In 100 ml of Conical flask, 0.32 g agarose powder was added and 40 ml of TAE (Tris acetate- EDTA) buffer was used to dissolve, oven-heated and cooled down. 4 µl EtBr (Ethidium Bromide) was added to the luke warm gel and poured on the tray to solidify the gel. The agarose gel was placed in the electrophoresis chamber and 3 µl of Genomic DNA samples and 2 µl of 100bp ladder were loaded to the well and run for 30 minutes to check the quality and concentration for Genomic DNA.

High quality Genomic DNA was used for Exome sequencing library preparation using Illumina Truseq Exome Enrichment Kit (Illumina). 100 ng of genomic DNA was fragmented using Covaris (ME220) Instrument, followed by end repair to get blunt end fragments. The enriched exome libraries were analyzed in

Agilent 2200 TapeStation using high sensitivity D1000 ScreenTape and the pooled libraries were loaded on S4 Flow cell (2 X 100 bp paired-end) sequencing run in Illumina Novaseq 6000 sequencer.

Sequencing of 15 tissue cancer samples and adjacent normal samples from triple-negative breast cancer (TNBC) patients and 15 blood (germline) samples were done. The raw sequencing data were processed using Fastp in paired-end mode, with a PHRED score cut-off of 30 used to remove low-quality reads and adaptors. FastQC was used to assess the quality of raw reads and trimmed reads that exceeded 30 base pairs and had an error rate below 10% were retained. Trimmed reads were aligned to the human reference genome GRCh38 using default parameters in BWA. Aligned reads were sorted and indexed with Samtools, and duplicate marking was performed using Picard. Base quality score recalibration was done using Genome Analysis Toolkit (GATK) with known sites vcf files of dbsnp version 146 and Mills and 1000G gold standard indels with default settings and Post-quality control.

Somatic variants were called using Mutect2 with matched tumor/normal pair mode, a variant caller within the GATK workflow. The output bam file was processed with gatk GetPileupSummaries with known variants, integrating population allele frequencies of common and rare alleles from gnomAD, alongside a bed file delineating exome intervals, effectively circumventing the calling of germline variants. Following this, the Mutect2-filtered variants were subjected to annotation via Ensembl Variant Effect Predictor (VEP). Further filtration based on contamination estimates was performed using GATK FilterMutectCalls, ensuring that only somatic variants meeting the filter criteria proceeded for subsequent analysis. The FILTER field is labelled with PASS for calls that are likely true positives, and 14 filters, including contamination, appear to be applied at this step.

The germline variants were called using GATK Haplotypecaller and annotated using Annovar. Following this, a manual filtering process, specifically focusing on exonic variants, non-synonymous and pathogenic variants, was carried out to select only the important variants. In order to further assess the functional impact of these variants, three different predictive tools, SIFT, PolyPhen2, and

MutationTaster were utilized. In the subsequent analysis phase, only exonic nonsynonymous variations deemed deleterious by at least two of the following tools: SIFT (labeling as Damaging), PolyPhen2 (labeling as Probably Damaging), and MutationTaster2 (labeling as Disease causing) were selected for downstream analysis.

The identification of mutually exclusive or co-occurring pairs of mutated genes within our samples was conducted using the somaticInteractions feature integrated within Maftools, which utilizes pairwise Fisher's exact test (to detect interactions among the top 30 most mutated genes). In cancer, numerous disease-causing genes exhibit either co-occurring patterns or distinct exclusivity in their mutation behaviors. To identify such sets of genes, the somaticInteractions function in Maftools is employed, employing a pairwise Fisher's exact test to pinpoint significantly associated gene pairs. Additionally, the somaticInteractions function incorporates cometExactTest to recognize potentially altered gene sets involving more than two genes. For this analysis, the top 50 driver genes were utilized, labelling the results with a threshold P value of 0.05 and 0.01. This process aids in unveiling exclusive or co-occurring relationships among mutated genes, contributing to a better understanding of their roles in cancer development and progression (Source: OncoVar: an integrated database and analysis platform for oncogenic driver variants in cancers).

The pathway analysis was performed using The Database for Annotation, Visualization and Integrated Discovery (DAVID) Bioinformatics database tool. This tool functioned by computing the proportions of genes affected within a pathway, as well as determining the fractions of samples displaying pathogenetic or unknown variants in these pathway genes.

Analysis of mutational signatures was conducted by utilizing our dataset sourced from filtered somatic variants in VCF files using Maftools in R package. The analysis method involved extracting neighboring bases adjacent to the mutated ones and creating a matrix for pathogenetic or unidentified variants. This matrix classified nucleotide substitutions into 96 classes based on the nearby bases,

achieved using the *trinucleotideMatrix* function. Non-negative matrix factorization (NMF) was then employed, utilizing the *estimateSignatures* function to estimate the number of signatures. Three optimal signatures were determined based on the cophenetic correlation metric. The pathogenetic or unknown variant matrix was further divided into these three signatures with the *extractSignatures* function. These extracted signatures were compared to both previously known COSMIC signatures—30 legacy signatures and a more recent set of 65 Single Base Substitution (SBS) signatures—using the *compareSignatures* function. The resulting signatures were then visually represented using the *plotSignatures* function.

RESULTS AND DISCUSSION

The study consisted of 59 cases, among them 39 (66.1%) cases > 45 years and 20 (33.9%) cases < 45 years. Chi-square goodness of fit test was used to compare the proportion of cases equal between the age group of cancer patients, frequency of exercise and sleep hrs. The results reveal that more cases were observed at the age group above 45 years (X2=6.119, P<0.05). Exercise, 40 (67.8%) of the cases did not do exercise and only 10 (16.9%) of cases did exercise every day (X2=31.559, P<0.05). The results show that 43 (72.9%) of the cases were having sleeping > 6 hrs and 27.1 % of the cases slept less than 6 hrs (X2=45.186, P<0.05).

High consumption of pork, fish, meat, oil and less consumption of fruits, vegetables and water are significantly associated with TNBC. Smoked meat and smoked vegetable consumption were not significant. Consumption of tobacco products/ alcohol were highly significant with TNBC. Different smoke, liquid form (tuibur) and smoke-less (Sahdah and khaini) forms of tobacco were practiced by the cases. Alcohol consumption also included the branded and local varieties.

Menarche, parity/ breast feeding/ miscarriage/ late menopausal age / familial cancer/ co-morbidities (diabetes and hypertension) / blood relatives with cancers are significantly associated with TNBC. These findings are made under the category of

reproductive history. IDC, tumor grade, Lymph node metastasis and DCIS are significantly associated with TNBC. BMI was found to be higher in younger age TNBC patients. Overall survival was higher in older age (> 45 years) TNBC patients. Consumption of Salt, Gutkha and tuibur are significantly associated with TNBC recurrence in our patient cohort. The biological parameters are not significantly associated with TNBC recurrence in our patient cohort. Other cancer types are not significantly associated with the above TNBC socio-demographic and biological parameters in our patient cohort. Other cancer types are significantly associated with sahdah, co-morbidities and relatives with ovarian cancer.

Comorbidities are not significantly associated with the above TNBC socio-demographic parameters in our patient cohort. Comorbidities are significantly associated with no. of children, age at first delivery, breast feeding duration and menopause. Familial cancers are significantly associated with bettlenut consumption. Familial cancers are significantly associated with first or second degree relatives with other types of cancer. Tumor type is significantly associated with water consumption in TNBC patients. Tumor type is significantly associated with khaini, age at menarche and breast-feeding duration in TNBC patients. Lymph node metastasis is significantly associated with age and comorbidities in TNBC patients. DCIS is significantly associated with Beef, fruits and sahdah consumption in TNBC patients. Salt consumption is significantly associated with survival in TNBC patients. All other demographic and biological parameters were insignificant.

Overall survival is significantly less in under-weight and over-weight in TNBC patients. For Asia-Pacific region, the BMI for normal is 18.5 to 22.9, Underweight is < 18.5 and Obese is ≥ 25. Exercise is mildly significant between the Healthy controls and TNBC patients. Regular consumption of Meat and smoked food were significant; Less intake of fruits and water are also significant between the Healthy controls and TNBC patients. Consumption of Tobacco products is also significant between the Healthy controls and TNBC patients. Breast feeding, Menopause and co-morbidities are significant between the Healthy controls and TNBC patients.

The most common of these missense mutations is a single-nucleotide polymorphism (SNP) mutation, which was characterized by a base change and faced multiple genetic variants. Furthermore, the most prevalent single nucleotide variation (SNV) is C > T mutations. The prevalence of C > T mutations suggests that most missense mutations are caused by cytosine (C) to thymine (T) changes at specific genomic loci, posing implications for understanding genetic mechanisms and disease associations. The number or variants per sample ranged from 4 to 131, with a median value of 37. Among the 10 most mutated genes, Maftools flags MUC16 as a gene that is usually found mutated in exome studies, therefore likely to be a passenger gene.

Several genes have been found to undergo mutations frequently in somatic TNBC. Among these, *TP53, CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD*, and *TLL1,* were the most frequently mutated genes. Moreover, *TP53* gene demonstrated the highest occurrence of genetic variants and short indels, present in 47% of the patients' tumour samples. Conversely, the remaining 15 genes collectively displayed a mutation frequency of 13%.

A comprehensive somatic interaction analysis was conducted to investigate the relationship between the most commonly mutated genes within a cohort of patients with somatic breast conditions. The primary focus was on ascertaining whether these mutations tended to appear together or exhibited a pattern of mutual exclusivity. Two pairs of genes, *OTULINL* and *CDH17,* as well as *MUC16* and *MT-CO3*, displayed a pronounced tendency to co-occur. This observation was supported by statistical significance, with a p-value of 0.05 for each gene pair. This implies a non-random, potentially interdependent relationship between these gene combinations within the context of somatic breast conditions. These findings may indicate the potential interactions influencing the development or progression of somatic breast conditions, suggesting possible cooperative mechanism or shared pathways between these gene pairs. Further investigation into the functional

implications of these gene interactions could provide valuable insights for targeted therapeutic approaches or diagnostic strategies in breast pathology.

A pathway analysis showed several genes associated with the oncogenic pathways. There were 22 genes associated with the PI3K-Akt signaling pathway. This pathway plays a crucial role in cell growth, survival, and various cellular processes. The MAPK signaling pathway, with 19 gene counts, is also of notable importance as it is involved in regulating cell proliferation and differentiation. Apoptosis, an essential mechanism for programmed cell death, had 13 genes associated with it. Furthermore, several pathways, such as the Rap1 signaling pathway, Phospholipase D signaling pathway, and Cellular senescence, had intermediate gene counts, suggesting their potential relevance in the study's biological context. In contrast, pathways with lower gene counts may still hold significance. For example, pathways like Starch and sucrose metabolism and Carbohydrate digestion and absorption had six gene counts, suggesting a potential role in metabolic processes.

A comprehensive analysis of drug-gene interactions has identified a number of genes with potential as druggable targets. These genes include *ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16,* MT-*CO3* and *TLL1*. Each of these genes underscores their suitability as targets for future drug development and therapeutic interventions.

In a germline variant analysis, 557 variants across the genome were identified, spanning 152 different genes after applying variant filtration to focus on the most clinically relevant ones. The variant filtration criteria were targeting exonic, non-synonymous and pathogenic variants only, resulting in a subset of 557 missense mutations. Furthermore, single-nucleotide polymorphism (SNP) mutations were the most frequently observed, followed by deletions. Notably, C > T mutations, categorized as pathogenic or of unknown significance, constituted the predominant class among single-nucleotide variants. The number or variants per sample ranged from 29 to 46, with a median value of 37 (Figure 16 D). Among the 10 most mutated

genes, Maftools flags MUC16 as a gene that is usually found mutated in exome studies, therefore likely to be a passenger gene.

The germline whole exome sequencing analysis of TNBC analysis identified a large number of genetic variants found in the germline blood samples, showing the predominant genes frequently altered. The most frequently mutated genes were *BRCA1, SUMO4, FANCA, BARD1, BRCA2, CCDC170, KLKB1, PRKCQ, COL4A4,* and *IL4R*. Among these genes, the prevalence of genetic variations varied from patient to patient. Notably, *BRCA1, SUMO4*, and *FANCA* displayed the highest prevalence of common genetic variants and short indels present in all patients (100%) from the tumor samples. Following this, *CCDC170* displayed a prevalence of 93% in the samples, while *KLKB1, BARD1, BRCA2, COL4A4, PRKCQ,* and *IL4R* displayed frequencies of 87%, 80%, 73%, 73%, 67%, and 67%, respectively.

Mutational signature analysis in germline TNBC revealed the presence of three distinct mutational signatures associated with 75 genes. Accordingly, the signature 1 consisted of 26 genes with a cosine similarity score of 0.608 demonstrate remarkable congruence with the COSMIC_6 mutational signature with an aetiology of defective DNA mismatch repair. Similarly, the signature 2 included other 26 genes show a cosine similarity score of 0.721 with the mutational signature of COSMIC_6 also of defective DNA mismatch repair aetiology. Furthermore, the signature 3 consisted of another subset of three genes match the COSMIC_1 mutational signature, scoring 0.713 on the cosine similarity scale. A defective DNA mismatch repair mechanism was found to be responsible for the etiology of these genes.

A pathway analysis revealed a variety of oncogeneic pathways associated with the genes including Pathways in cancer, PI3K-Akt signaling pathway, Human papilloma virus infection, Platinum drug resistance, Hypertrophic cardiomyopathy, Gastric cancer, Breast cancer, Colorectal cancer, FoxO signaling pathway, Mismatch repair, Central carbon metabolism in cancer, PD-L1 expression and PD-1 check point pathway in cancer and Thyroid cancer.

Drug–gene interactions analysis in germline samples identified eighteen genes as potential druggable targets, including *BRCA1, NQO1, BRIP1, IL4R, COL4A4, FANCA, TTN, F5, APOB, TGFB1, KLKB1, BRCA2, ARSA, SERPINA7, BARD1, PRKCQ, GPSM2* and *MSH6.* These genes are potential targets for future drug development and therapeutic interventions. A rich assortment of genes offers opportunities for further research and possible breakthroughs in precision medicine, promising tailored treatments and a better understanding of their role in physiological and pathological processes.

Pathological and genetic investigations are the gold standard for accurate diagnosis of cancers. This study will try to identify several genes that associate to TNBC and its association with risks and overall breast cancer for improved cancer risk assessment. The information on gene-specific risks for TNBC is useful to identify the genes. This study is important as it aims to characterize the immunological profile and mutations. It will also help to find out the pervasiveness of mutations unique to Mizo ethnic groups and TNBC. This may help in better clinical management of individuals at risk for or diagnosed with the cancer.

In Mizo population, about 67.96% of the cases had breast cancer when they were older than 40. Exercise, sleeping patterns, or night work do not appear to be significantly correlated with tumor grade according these studies. Tumor grade may be influenced by the frequency or amount of consumption of any particular variable, including pig, fish, beef, chicken, fruits, vegetables, Sa-um (fermented pork fat), smoked meat, smoked vegetables, oil, salt, and water. An analysis of the relationship between the consumption of alcohol and tobacco and tumor grade revealed no significant results for any of the variables.

With 65 cases out of 129 control samples, the age range of 41 to 51 years has the highest risk factors. Obese women may have a higher than average chance of developing breast cancer. A significant p-value of 0.000 was found for BMI > 30, indicating that 42.3% of the cases were obese. Increased breast cancer cases are significantly correlated with higher BMI. In contrast, family history is quite significant. 5.2% of respondents who

were control subjects have a history of breast cancer, and only 5.6% of them were married.

About 80% of cases of breast cancer, including TNBCs, occur in people older than 50. Risk increases to 1.5% at age 40, 3% at age 50, and over 4% at age 70 years (Almansour, 2022). Women's breast cells are especially susceptible to hormonal abnormalities including the hormones progesterone and estrogen. The circulation of androgens and estrogen has been associated to a higher risk of breast cancer. The profile of a BRCA2-associated tumor is correlated with luminal-like breast cancers, namely the Luminal B subtype, whereas that of BRCA1-related tumors is similar to the TNBC subtype. TNBC is still very common in non-Hispanic white women. Additionally, black women are thought to have the lowest rates of cancer survival, and their death rate is substantially greater. Like TNBC, a major risk factor for breast cancer is one's family history**.** A first-degree breast cancer relative is reported by 13–19% of patients with breast cancer diagnoses.

Compared to postmenopausal and white women, African American and premenopausal women had a significantly greater chance of developing basal-like BC and a significantly reduced chance of developing luminal A BC. The study population had a higher TNBC prevalence than white BC patients. TNBC was independently correlated with younger age, premenopausal status, increased parity, use of hormonal contraceptives, high histological grade, and advanced disease. The body mass index (BMI) and HR expression were inversely correlated in premenopausal women. On the other hand, BMI positively correlated with HR and HER2 levels in postmenopausal women.

This study investigated into the comprehensive analysis of whole exome sequencing data sourced from 15 patients diagnosed with triple-negative breast cancer in Mizoram. The dataset encompassed 15 tissue cancer samples, corresponding adjacent normal tissues, and 15 blood (germline) samples. We initiated our study by performing Whole Exome Sequencing (WES) analysis on somatic tissue samples of Triple-Negative Breast Cancer (TNBC). Subsequently, we investigated into germline variants analysis. These comprehensive analyses encompassed Single Nucleotide Variant analysis, identification of frequently mutated genes, exploration of mutually exclusive gene patterns, deciphering

mutational signatures, assessing drug-gene interactions, and analysing pathways involved. Furthermore, our investigation extended to clinical data analysis pertaining to somatic TNBC samples, leading to the discovery of single nucleotide variants (SNVs) and their related genes associated with TNBC.

WES analysis of the somatic TNBC samples identified a total of 188 somatic single nucleotide variants (SNVs) and revealed associations with 647 genes. These variants exhibited diverse types, with the most prevalent being nonsense mutations (560), followed by frame shift mutations (37) and several other mutation categories. The germline WES analysis in blood samples unveiled 557 variants within 152 genes after filtering for exon, non-synonymous mutations and pathogenic variants only. Notably, among the identified variants, amino acid substitutions in proteins, largely attributed to genetic disorders, were predominantly linked to missense mutations arising from single nucleotide variations in both cases. This observation highlights the significance of these diverse mutations in influencing genetic anomalies, with missense mutations, in particular, holding a substantial role in protein-level alterations and consequent biological implications. The comprehensive breakdown of these findings unveils a significant understanding of the genetic landscape within the context of potential genetic disorders and offers a basis for further investigation into their functional and clinical implications.

This analysis of somatic variants unveiled TP53 as the predominant gene exhibiting a notably high frequency of genetic variations within the tumor samples studied. Notably, 47% of the patients displayed mutations in TP53 that signifies the prevalence of TP53 mutations in somatic breast cancer patients, whereas the remaining 15 genes collectively showcased a 13% mutation frequency, signifying their relevance in the landscape of genetic alterations observed in these samples. The diverse array of genes, including *CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD*, and *TLL1,* collectively contribute to the mutational landscape. On the other hand, the most frequently mutated genes in germline samples are *BRCA1, SUMO4*, and *FANCA* that displayed the highest prevalence of common genetic variants and short

indels present in all patients (100%). Following this, *CCDC170* displayed a prevalence of 93% in the samples, while *KLKB1, BARD1, BRCA2, COL4A4, PRKCQ,* and *IL4R* displayed frequencies of 87%, 80%, 73%, 73%, 67%, and 67%, respectively. These findings are consistent with previous studies, underscoring the consistency and relevance of these genetic variations in breast cancer research.

Additionally, significant co-occurrences were identified between two pairs of genes in somatic TNBC samples *OTULINL/CDH17* and *MUC16/MT-CO3*, signifying associations with p-values of 0.05 whereas, none of the genes are mutually exclusive. On the other hand, in germline samples, a significant co-occurrence tendency was observed between *PRKCQ-BARD1* and *APOB-NQO1*. The presence of these pairs suggests a possible interaction or interdependence between these two pairs indicating that they may be interdependent or mutually associated in TNBC. Mutations in one gene may increase the probability or impact of mutations in another, suggesting that there may be synergistic effects or shared pathways. Additionally, six gene pairs, such as *F5-ARSA, MTHFR-GHRL, F5-BRIP1, TTN-GPSM2, APOB-MTHFR* and *GPSM2-IL4R*, appeared to exhibit a contrasting pattern by being mutually exclusive, suggesting that these genes tend not to occur together in the same germline TNBC samples. For instance, in cancer, mutations in different genes may be mutually exclusive, which implies that their functions are similar within one pathway, whereas mutations in one gene might prevent mutations in another gene.

Furthermore, the mutational signature analysis in the somatic TNBC samples identified three distinctive signatures associated with TP53 gene. Signature 1 and 2 match COSMIC_6 mutational signature with an aetiology of defective DNA mismatch repair. DNA mismatch repair (MMR) is responsible for preventing mutations and genetic instability by identifying and repairing mismatches or small distortions that may occur during DNA replication. The signature 3 closely match the COSMIC_3 mutational signature with an aetiology of defects in DNA−DSB repair by HR (pathway to repair DNA double-strand breaks). These DNA repair problems could contribute to genetic instability and cancer-related diseases (Aparicio

et al., 2014). On the other hand, the mutational signature analysis in the germline TNBC samples revealed the presence of three distinct mutational signatures associated with 75 genes. Signature 1 consisted of 26 genes and Signature 2 consisted of another 26 genes, both of them congruent with the COSMIC_6 mutational signature with a defective DNA mismatch repair aetiology. In addition, signature 3 consisted of another subset of three genes, which matched the COSMIC_1 mutational signature and had an etiology of a defective DNA mismatch repair mechanism. DNA mismatch repair (MMR) is responsible for preventing mutations and genetic instability by identifying and repairing mismatches or small distortions that may occur during DNA replication.

Pathway analysis in somatic samples showed several genes associated with the oncogenic pathways including PI3K-Akt signalling pathway, MAPK signalling pathway, Apoptosis, Rap1 signalling pathway, Phospholipase D signalling pathway, and Cellular senescence. Whereas, in the oncogenic pathways identified in germline samples include Pathways in cancer, PI3K-Akt signalling pathway, Human papilloma virus infection, Platinum drug resistance, Hypertrophic cardiomyopathy, Gastric cancer, Breast cancer, Colorectal cancer, FoxO signalling pathway, Mismatch repair, Central carbon metabolism in cancer, PD-L1 expression and PD-1 check point pathway in cancer and Thyroid cancer.

Drug-gene interactions have been extensively investigated, and thirteen genes have been identified as potentially druggable targets for somatic samples, including *ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16, MT-CO3* and *TLL1*. In germline samples, eighteen genes have been identified as possible druggable targets, including *BRCA1, NQO1, BRIP1, IL4R, COL4A4, FANCA, TTN, F5, APOB, TGFB1, KLKB1, BRCA2, ARSA, SERPINA7, BARD1, PRKCQ, GPSM2* and *MSH6*. These genes are frequently mutated genes and may potentially targets for future drug development and therapeutic interventions. A rich assortment of genes offers opportunities for further research and possible breakthroughs in precision medicine, promising tailored

treatments and a better understanding of their role in physiological and pathological processes

The analysis of clinical features along with the somatic TNBC samples identified several variants and their associated genes including 39 variants associated with Age at menarche, 36 variants linked to co-morbidities, and another 36 variants concerning the distinction between the left or right breast cancer. Additionally, 34 variants were correlated with salt intake, while 33 variants were found associated with Location, and 30 variants with Grade, signifying their prominence among the factors examined. Nevertheless, within this cohort study, the analysis did not indicate any significant correlations between Alcohol intake, Age at menopause, and First-degree family history of breast cancer, suggesting a lack of substantial association within this specific dataset. In contrast, since there are only 15 samples, some clinical characteristics may not be sufficient to make accurate predictions due to the limited sample size. However, the study of clinical attributes and their genomic variants provide a clear understanding of how genetic variants are interconnected with distinct clinical features, revealing the significant impact of certain genetic elements on various aspects of breast cancer-related characteristics.

Cancer remains a global health challenge, necessitating continuous research to unravel its intricacies and develop effective therapeutic strategies. One fundamental aspect of cancer etiology is the accumulation of genetic mutations, both in somatic and germline cells, which contribute to the initiation and progression of malignancies. In this study, we provide a comprehensive account of the results obtained from an in-depth analysis of somatic mutations, germline variants, mutational signatures, and their impact on crucial oncogenic pathways. We explore how genetic diversity contributes to the complexity of this disease and integrate clinical data to identify genes or genetic variations linked to clinical outcomes. We also integrate clinical data to identify genes or variants related to the particular effect of the clinical causes. We defined the genetic alterations in 15 triple negative breast cancers using a whole exome sequencing data with a clinical data analysis including somatic and germline data from Mizoram, India. The most frequent somatic

alterations were mutations in *TP53* (47%) and *ELAPOR1* (13%) and *CACNA1E* (13%) and other genes at 13% mutation rates among the top 10 mutated genes. Germline mutation revealed that among top mutated genes are *BRCA1* (100%), *SUMO4* (100%), *FANCA1* (100%), *CCDC170* (93%), *BARD1* (80%) *BRCA2* (73%). The mutational signature analysis revealed the presence of three distinct mutational signatures associated with 20 genes including the etiology of a defective DNA mismatch repair mechanism in COSMIC database. Furthermore, our pathway analysis revealed that it involves somatic variants include many oncogenic pathways are altered including PI3K-Akt signaling pathway, MAPK signaling pathway, Apoptosis and Rap1 signaling pathway these are the main altering pathway for carcinogenesis. This study represent the first whole exome exome sequence data of somatic and germline mutation analysis from Mizoram triple negative breast cancer patient.

SUMMARY

- Somatic cancer analysis identified a total of 688 variants spanning 647 genes and the most frequently mutated genes are *TP53, CACNA1E, ELAPOR1, CDH17, CSMD1, IGSF3, MT-CO3, MUC16, NPIPB15, OTULINL, PIK3CA, RYR1, SNAI3, SYNE2, THBD,* and *TLL1*.
- *TP53* exhibited the highest mutation rate, exceeding 43%, while the other genes had mutations around 7%.
- These findings highlight the significant role of *TP53* mutations in breast cancer among these patients.
- Germline variant analysis identified 557 variants in 152 genes after filtering for exon and non-synonymous mutations.
- *BRCA1, FANCA*, and *SUMO4* mutations were found in all blood samples.
- Risk factors for TNBC include young age at breast cancer diagnosis, young age at menarche, young age at time of first child birth, high parity, lack or shorter duration of breast feeding, premenopausal women with high body

mass index and African American ethnicity and an elevated waist: hip ratio (Dawood, 2010).

- Some of the risk factors for TNBC are BRCA mutation, ethnicity, age and Body Mass Index (BMI) (Mousavi et al., 2019).
- TNBC has a worse prognosis, high recurrence and poor survival rates when compared to other subtypes of breast cancer (Dawood, 2010).

- Some of the risk factors for TNBC are BRCA mutation, ethnicity, age and Body Mass Index (BMI) (Mousavi et al., 2019).
- TNBC has a worse prognosis, high recurrence and poor survival rates when compared to other subtypes of breast cancer (Dawood, 2010).

- PALB2 has been identified as one of the common predisposing genes for breast cancer afer BRCA1/2 with penetrance estimated at 33–70% depending on age at diagnosis and family history (Siraj et al. 2023).
- BRCA1 and BRCA2 germline mutations are responsible for about 60% of HBCs with an overall 60–80% lifetime risk; the other 40% are associated with other predisposing variants in moderate-to-high penetrance genes such as PALB2, PTEN, TP53, CDH1, CHEK2, ATM, and the MMR group (Catana et al., 2023).
- 32.8% patients carried a pathogenic or likely pathogenic heterozygous germline mutation in 19 genes. including high-penetrant breast cancer genes like BRCA1 (43%), BRCA2 (19%), PALB2 (12%), TP53 (3%), and CDH1 (2%) (Catana et al., 2023).
- Eight genes were significantly mutated in the Breast cancer tumors including 54 PIK3CA, TP53, GATA3, MAP3K1, CDH1, CBFB, PTEN, and RUNX1 (Ding et al., 2023)
- EGFR pathway contains well-established oncogenes including EGFR, KRAS, BRAF and PIK3CA genes that modulate gene activations in solid

tumors including lung, colorectal cancer (CRC) and pancreatic ductal adenocarcinoma (Zakaria et al. 2019).

- In our study, Significant gene co-occurrences were observed, including *OTULINL/CDH17* and *MUC16/MT-CO3,* with p-values of 0.05.
- Mutational signature analysis revealed three distinct signatures associated with 20 genes, with SBS 6 indicating defective DNA mismatch repair.
- Drug-gene interaction analysis identified potential druggable genes, including *ALDH3B2, TP53, IGSF3, SNAI3, THBD, ABCA13, CDH17, RYR1, CACNA1E, PIK3CA, MUC16, MT-CO3*, and *TLL1*.
- Clinical data analysis identified key factors contributing to the cohort's outcomes, such as age at menarche, comorbidities, breast location, salt intake, and grade. However, no significant associations were found for alcohol intake, age at menopause, or first-degree family history of breast cancer.