# CHARACTERIZATION OF CLINICALLY SIGNIFICANT MUTATIONS ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

**DORIS ZODINPUII**

**DEPARTMENT OF BIOTECHNOLOGY**

**MIZORAM UNIVERISTY**

# Characterization of clinically significant mutations associated with Breast Cancer in Mizo population

**By**

**DORIS ZODINPUII**

**Department of Biotechnology**

**Submitted**

**In partial fulfillment of the requirement of the Degree of Doctor of Philosophy**

**in Biotechnology of Mizoram University, Aizawl**

**Department of Biotechnology**
School of Life Sciences
**MIZORAM UNIVERSITY**
(*A Central University*)
Tanhril**,** Aizawl - 796 004, Mizoram

**Dr. N. Senthil Kumar** Ph.D
Professor

Phone: 09436352574 (M)
E-mail: nskmzu@gmail.com

## CERTIFICATE

This is to certify that the thesis entitled **"Characterization of clinically significant mutations associated with Breast Cancer in Mizo population"** to Mizoram University for the award of the degree of Doctor of Philosophy Biotechnology by **Doris Zodinpuii** Registration No. **MZU/Ph.D/939 of 15.11.2016**, Ph.D scholar in the Department of Biotechnology, under my guidance and supervision and has not been previously submitted for the award of any degree in any Indian or foreign University. She has fulfilled all criteria prescribed by the UGC (Minimum Standard and Procedure governing Ph.D. Regulations). She has fulfilled the mandatory publication (Publication enclosed) and completed Ph.D. course work.

This also certifies that the scholar has been admitted in the Department through an entrance test, followed by an interview as per clause 9(i) and (ii) of the UGC Regulation 2009.

Place: **(Prof. N. Senthil Kumar)**

Date: **Supervisor**

## Declaration

I, **Doris Zodinpuii**, hereby declare that the subject matter of this thesis entitled "**Characterization of clinically significant mutations associated with Breast Cancer in Mizo population**" 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 do 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.

Date:

(Doris Zodinpuii)
**Candidate**

(Dr. Th. Robert Singh)
**Head**

(Prof. N. Senthil Kumar)
**Supervisor**

**Acknowledgement**

Above all, I am grateful to the almighty **God** for his unfailing guidance that had lead me to the end of my work.

I wish to express my heartfelt gratitude to my supervisor **Prof. N. Senthil Kumar**, for giving me an opportunity to be a part of his research group, for his constant support, guidance, encouragement and patience throughout my work.

I am also thankful to **Dr. Th. Robert Singh**, Head, Department of Biotechnology and all the **faculties** and **technical staff** for their help, support and kind cooperation.

I would like to express my deepest gratitude to all the participants and all the collaborating doctors, **Dr. Jerry L.Pautu**, **Dr. B. Zothankima** and all the staffs of MSCI for their kind support to access the patients and their medical reports, without whom, this work would not have been possible.

I owe my gratitude to **Dr. Pradnya Kotwal** and **Dr. Ashok Varma** for helping me countless times in many ways during my course of work.

I would also like to acknowledge the Advanced Level State Biotech Hub (BT/04/NE/2009) facility and Bio-informatics Infrastructure Facility (No.BT/BI/12/060/2012(NER-BIF-MUA), sponsored by Department of Biotechnology (DBT), New Delhi, Govt. of India for providing infrastructure to carry out this work to a success.

To all my lab mates and research scholars at Department of Biotechnology for their support.

Last but not the least, I would like to thank my parents **R.Zohmingthanga** and **Laltanpuii**, all my family, my beloved husband **Lalrinfela** and my son **Daniel Malsawmkima** for their contant support, love and prayers.

Date: Doris Zodinpuii

## 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 |
| $^{o}$C | -Degree Celcius |
| 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 |

## Table of Contents

## Lists of Figures

## Lists of Tables

# Chapter I

## Introduction and Review of Literature

## *Introduction*

The word Cancer is used to refer cluster of many different associated diseases and it can occur in almost the entire body. In every cancer, the cells, which are the basic units of human body begin to divide without stopping and form a clump of cells referred to as tumor. A tumor is not always cancerous, but can be benign as well when the cell grows but does not spread. A tumor is said to be cancerous (malignant) when they have the ability to multiply and propagate to any other location in the body. All cancers do not form a solid tumor and those are the cancers of the blood.

Cancers are of different types and are normally named after the organ or tissues they have originated. For example: Breast Cancer (BC), Prostate Cancer and Lung Cancer start from the cells of breast, prostate and lung, respectively. The word Carcinoma refers to the cancers of epithelial cells that are found lining the interior and exterior boy surfaces. Of all the cancer, the occurrence of carcinoma is most common. Further they are group into distinctive types of cancer based on the epithelial cell involved. Cancers of fluids or mucus producing epithelial cells are called Adenocarcinoma. Basal cell carcinomas are those that begin in basal layer, which is the lower region within the epidermis of the skin. Cancer that is formed by the epithelial cells above the basal layer is called Squamous cell carcinoma. They are sometimes referred to as epidermoid carcinomas. Lastly, cancer that is forms in the transitional epithelium or urothelium is called Transitional cell carcinoma. The lining of the ureters, bladder and a small region of the kidneys (renal pelvis) consist of these tissues and cancers of these organs could be transitional cell carcinomas (National Institute of Health, 2019).

**Breast Cancer**

The breast is located in the chest pectoral muscles. Female breasts are mainly composed of lobules, ducts and connective tissues. The lobules are a small structured gland that produces milk. A cluster of lobules is called the lobes. These lobes organized the breast into 15-20 sections. The duct is a network of tube or stem like structure. The duct is connected to the lobule and supplies the nipples with the milk. The area surrounding the nipple with a darker skin is called the areola. The connective tissue supports and maintains the architecture of the breast. They are mainly comprised of fat cells (fibrous and adipose tissue). The adipose tissue spread all over the collarbone, under the arm and across to the middle of the ribcage. Amount of fat content is responsible for the breast size (Human Anatomy, 2019; Centers for disease control prevention, 2019).

A malignant tumor initiating from the breast cells, usually in breast lobules or ducts is called BC. On a rare occasion, BC can also develop in the connective tissues. This cancer cells are said to metastasize when they travel through the lymph node and invade near and distant healthy tissues of the breast and other location in the body. BC staging is based on the distance the cancer cells have spread from the tumor origin (Breastcancer.org, 2019).

**Types of Breast Cancer**

Generally, BC is carcinomas, also referred in a more specific term as Adenocarcinoma. Depending on the severity and site of cancer origin, BCs are of

several types. If the cancer remains in the original site, they are referred as "In situ" and the term "Invasive or infiltrating" is used to represent when the cancers had invaded the nearby tissues. The name of the tissues or cells used to show the location of cancer initiation. For example, it could be a ductal or lobular carcinoma. They are: -

**1. Ductal carcinoma in situ (DCIS)**: They are cancers formed in the ductal cells of breast. DCIS is non-invasive since the cancer remains in the milk duct and does not extent in the neighboring breast tissue. DCIS is easy to manage, but once developed it increases the chance of developing into invasive BC later on. Also, having a history of DCIS increases the risk of BC recurrence by 30% than those who have not had a DCIS.

**2. Invasive/ Infiltrating Ductal Carcinoma (IDC)**: IDC is observed in approximately 80% of all BC cases diagnosed. In IDC, the nearby healthy breast cells are invade by breaking the lining of the milk duct by the cancer cells. IDC is also found in men but most commonly observed in older women. IDC is further divided into five subtypes, they are: -

i). *Tubular carcinoma*: Tumors are usually smaller in size and are made up of tubules that are tube-shaped structures. The tumors are often low-grade, and with the advancement of mammography, tubular carcinoma is thought to account for about 27% of all BC. It is commonly diagnosed in women of early fifties but is rarely diagnosed in men. Tubular carcinomas often tend to be less aggressive, respond well to treatment.

ii). *Medullary Carcinoma*: It represents approximately 3-5% of all BC cases. The

tumor resembles the delicate flesh of brain called medulla and is so called the "medullary" carcinoma. They are frequently observed in BC of age 40 to 50 years and also in patients with *BRCA1* mutation. It is commonly diagnosed among the Japanese women when compared to the women in United States. The tumor appears to be of high grade but behave as a low-grade tumor. The tumor grows slow and treating is easier than any other types of BC.

iii). *Mucinous Carcinoma*: Among the uncommon subtype of IDC is Mucinous carcinoma, also sometimes referred to as colloid carcinoma. The abnormal tumor cells are found floating in pool of slimy and slippery substance called mucin, which is a key ingredient of mucus. About 5% of invasive BCs consist a mucinous element along with other cancer cell type. It is rarely found in men and could be found at any age, 60s or early 70s is the average age at diagnosis. They are not aggressive and respond better to treatment.

iv). *Papillary Carcinoma*: In invasive BCs, papillary carcinomas account for less than 1-2% and is mostly found in postmenopausal elder women. A well-distinct border, made up of tiny projection that looks like finger, usually characterizes it. Invasive papillary carcinoma is often grade 2 or moderate grade.

v). *Cribiform Carcinoma*: In cribiform carcinoma, the stroma, which is the connective tissues of the breast found in between the ducts and lobules are invaded. Well-defined holes, which look a Swiss cheese, are formed in the tumor cells. The tumor is usually low grade. Cribriform represent only a tiny portion of tumors in less than 6% of invasive BCs.

**3. Lobular Carcinoma In situ (LCIS)**: Originated in the lobules, LCIS is an

abnormal growth of cell in an area that are non-invasive. Even though the names include the term carcinoma, LCIS is not an actual BC. Rather, having LCIS indicates that a person has a greater chance of developing BC. Due to this reason, the term, lobular neoplasia (collection of abnormal cells) is preferred by some experts, rather than lobular carcinoma. It is usually diagnosed in a premenopausal, mostly of ages 40 and 50 years.

**4. Invasive/ Infiltrating Lobular Carcinoma (ILC)**: It is the 2nd most prevalent BC type observed after IDC. The tumor starts in the lobule and invades the surrounding tissues. ILC accounts for about 10% of all invasive BCs. Even though ILC can develop at any age, the prevalence tends to increase with increased in age. As per the American Cancer Society, 2/3 of women diagnosed with ILC are older than 55 years of age. It has also been suggested using hormone replacement therapy could increase the chance of developing ILC.

**5. Inflammatory BC (IBC)**: Inflammatory BC is a rare, aggressive and rapidly growing cancer. It accounts for only 1% of all BC cases in United States. Study conducted in 2008 found that being overweight increases the risk of developing IBC.

**6. Metastatic BC**: When BC travels and invade different sites of the body, more frequently to the liver, brain, bones or lungs, they are called Metastatic BC or stage IV BC. BC cells can break their way out from their original site and move to a different location via. the lymphatic system or the bloodstream. Recurrence rate is higher and 30% of women diagnosed with early stage BC are believed to develop a metastatic disease.

**7. Paget's Disease of the Nipple**: In this case, a malignant rash developed in or

around the nipple's skin or areola. It is rarely found and account for only one percent of all BCs. Usually, the cancer first affects the milk ducts, then the surface of the nipple and finally the whole areola. Texture of the affected area becomes red, scaly, irritated and itchy. Around 97% of Paget's disease infected patients either have invasive cancer or DCIS elsewhere in the breast. Signs and symptoms of Paget's disease are often an alarming indication that BC have initiated.

**8. Phyllodes Tumors**: Phyllodes tumors, also known as Cystosarcoma are rare and account for a very less percentage (< 1%). Tumors often grow rapidly, but rarely spread out of the breast. Phyllodes tumors are form in the connective tissue of the breast rather than the duct or lobules. Majority of the phyllodes are benign and a much lesser percentage are malignant.

BC is also grouped into 4 intrinsic or molecular subtypes depending on the hormone receptors namely estrogen (ER) and progesterone (PR) and HER2/neu oncoprotein expression. They are: -

1. **Luminal A**: The cancer cells showed an expression of hormone receptors (ER+/ PR+) but not the HER2/ neu oncoprotein (HER2-). The protein that helps regulate cancer cells growth called "Ki-67" level is also low. The tumors are of low grade, tend to grow slow with better prognosis.
2. **Luminal B**: In luminal B, there is an expression of ER+ and PR+ whereas, Her2/ neu oncoprotein are either positive or negative with higher Ki-67 levels. Prognosis is poorer since the cancer usually grows at a higher rate than that of luminal A.
3. **Triple-negative/basal-like**: In case of triple negative BC, there is no expression

of either the hormone receptor or the HER2/ neu oncoprotein (ER-/PR-/HER2-). Triple negative BC is frequently observed along with *BRCA1* gene mutations and also in younger women.

4. **HER2-enriched**: HER2/neu oncoprotein expressions are enriched but not the hormone-receptors. HER2-enriched cancers, shows a faster growth rate with a poor prognosis. HER2 targeted therapies that include Tykerb (lapatinib), Perjeta (pertuzumab), Herceptin (trastuzumab), Kadcyla (T-DM1/ ado-trastuzumab emtansine) and Nerlynx (neratinib) are often used to treat this kind of BC. (Breast cancer. Org, 2019; American cancer society, 2019; Kumar et al., 2010)

**Signs, Symptoms and Diagnosis of BC**

Any of the following listed complications or discomforts could be the signs or symptoms of BC: -

1. Thick and Stiffened tissues in any area of the breast.
2. Swelling of all the breast or specific area.
3. Sensible lump in the breast or in the area of armpit.
4. Pain breast or armpits that is not altered by monthly cycle.
5. Eczema like rash on or around the nipples.
6. Nipple discharges other than milk, which sometimes contains blood.
7. Deformed nipple that has either inverted or sunken.
8. Change of either shape or size of the breast.
9. Dry, flaky and scaly skin of the nipple or breast.
10. Unusual texture (indented or dimpling) or color (redness) of the breast skin, like

an orange peel.

BC could be diagnosed by biopsy where a small tissue sample is surgically obtained to see if there are cancer cells. This procedure could tell the type of cancer and their hormone sensitivity. Routine screening could also be performed using an imaging test such as mammography (type of X-ray), ultrasound and Magnetic resonance imaging (MRI). These techniques could provide images of any lumps present in the breast, differentiate solid mass from cyst containing fluid and also information on how far cancer has reached (Breast Cancer.org, 2019). BC can also be inherited as well as sporadic.

1. **Inherited BC**: It accounts for a smaller number (5-10%) of BC cases and is caused by the autosomal inheritance of BC susceptibility genes mutations i.e. *BRCA1* and *BRCA2*. Mutation in these two genes account for only 10-20% of inherited BC. Mutation in other genes (*TP53, ATM, PTEN, CHD1, CHEK2, SKT11, BRIP1, PALB2,* etc) has also been reported to increase the susceptibility of developing BC. Inherited BC is generally characterized by onset of cancer at a younger age (~40 years) (Campeau et al., 2008; Jian et al., 2017; Larsen et al., 2014).
2. **Sporadic BC**: Sporadic BC accounts for about 90% of all cases. Besides those hypothesized low-to-moderate penetrance or modifier genetic alleles conferring a slight increase risk, several reproductive, environmental and demographic factors contributes a vital role in the development of BC. The onset is late in age and the cancer is commonly unilateral (Miki et al., 1994; Sirisena et al., 2018).

**Review of Literature**

BC is a complex multifactorial disease regulated by the association of various genetic, hormonal and environmental factors including diet, lifestyle and reproductive history. Globally, BC is the top cancer and causes of death among women (Atif et al., 2018; Sirisena et al., 2018). BC related death of around 327,000 has been estimated every year. Approximately 1.35 million new cases of BC have been diagnosed each year while 4.4 million women are believed to live with BC worldwide. Around 1.7 million women have been estimated to be diagnosed with BC in 2020 (Wong et al., 2009). In 2018, new cases of BC, along with lung cancer shares the same highest position (11.6%) when both the sexes are considered. And when only females are considered, the number of new cases (24.2%), the estimated incidence and mortality rate is also highest for BC (Globocan iarc, 2018).

In India, almost 100,000 cases of BC have been diagnosed each year, and an increased to 131,000 cases is expected by the year 2020 (Agarwal et al., 2008; Mangtani et al., 2010). According to the latest published of Globocan (2018), of all the new cases of cancer diagnosed in India, BC case is the highest when both sexes are considered and when only females are considered separately. Following the global trend, BC also top the age standardized incidence rate (24.7%) and mortality rate (13.4%) (Cancerindia.org, 2018). In the North Eastern part of India, the prevalence of BC is considerably higher then the rest of India probably due to the tobacco exposure leading to genotoxic stress (Saxena et al., 2010). Mizoram, located in the Southern corner of Northeast India, also shows an alarming trend in cancer. Nevertheless, BC threat remains the same in Mizoram. Latest report of National Cancer Registry Program (NCRP) revealed consistently high incidence rate of cervix

uteri (15.9 %), lung (15.6 %), and BC (13.0 %) among the Mizo women populace. BC incidence rate is believed to be already double by 2018 since the recent NCRP is not yet available. Hence, the rapid rise in the number of incidence and mortality because of BC among women in rural area like Mizoram (not to mention the whole world) over the past few year calls for an extensive study and exploration on this area of cancer (PBCR, 2014).

### *1. Epidemiological studies on BC*:

Variations in the incidence and mortality rates of BC between races or ethnicity of different parts of the world suggests that the known risk factors for BC may differ depending on the demographic and environmental factors and could not be explained by genetics alone (Lodha et al., 2011). Several factors that have been reported to help in BC progression are:

**A. Reproductive Factors**: It is well established that the development of BC is triggered the most by reproductive and hormonal factors. Reproductive factors contribution for BC development in Indian and western population is far different because of the differences in parity, early first delivery age and lactation which are part of Indian culture and less prevalent in western women. In developed countries, late age at first delivery, nulliparity, and lack of lactation are the leading reproductive risk factors (Palachandra et al., 2017).

Pregnancy provides protective effect by causing a permanent change of the breast cells and making them less prone to carcinogenic factors. This mechanism include differentiation of the breast epithelial cells, reduction in numbers of breast stem cells, modify breast response to estrogen and reduced the levels of circulating hormones (Opdahl et al., 2011).

Lactation has been observed to have a strong association among BC patient from North India. Among premenopausal women, lifetime duration of lactation was inversely associated with increased chance of BC development (Malvia et al., 2017). A landmark studies involving 50,000 BC patients from 47 epidemiologic studies have found that among parous women, there is a 4.3% reduced BC risk for every twelve months a woman lactates and a 7% reduced risk for every child birth independently. Following pregnancy, breastfeeding assist breast cells differentiation, and differentiated cells are rarely cancerous. Also, cessation process of breastfeeding (e.g., apoptosis) remove initial damaged DNA from the tissue of breast and is believed to reduced BC risk (Anstey et al., 2017).

Extended exposure to ovarian hormones in terms of menarche at an early age and late menopausal is found to have convincing association with increase BC occurrence. These ovarian hormones (estrogen and progesterone) regulate the breast mitotic activity and determine the possibility of tumorigenic somatic events (Singh et al., 2011; He et al., 2012). Usage of oral contraceptive pills confers a 5 times higher chance of BC development compared to non-user. A study on contraceptive pills used in association with BC involving 53,297 BC patients have found slight increase risk on women having contraceptive pills for 10 years, and also showed that there is

a 24% increased risk of BC among contraceptive pills compared to those who never used it (Alghamdi et al., 2015).

Early age at marriage is found to have a strong link with BC among Saudi females, and females getting married before age 18 have been found to have a 13.9 times higher risk of getting the disease. The reason for the increase risk may be explained by the inconsistency or elevation in the secretion of ovarian hormones, for example; hormone estrogen among young females (Alghamdi et al., 2015). Less number of parity or children is linked with the lifestyle of South Indian and Western women and has been found to increase BC risk. 40-50% risk reduction is associated with 3 or more number of live births. Single or unmarried women have elevated risks for developing BC and are usually late stage (Malvia et al., 2017).

**B. Environmental factors**

*Lifestyle***:** Physical activity, sleep pattern and night shift work have been shown to have an influence in BC development. The World Cancer Research Fund/American Institute for Cancer Research (2007) states that, there has been significant epidemiologic evidence supporting the inverse relationship between BC and physical activity in post-menopausal women. Several studies have supported that the risk of BC is approximately 25% lower among physically active compared to those inactive women (Hildebrand et al., 2013). Reduced risk is also observed on women who spent more time on household chores. Decrease in sex hormones and adiposity, increase immune functions and reduced markers of insulin resistance have been proposed to be the underlying biological mechanism by which physical activity protect against

BC (Malvia et al., 2017). With societal development, sleep pattern and night shift work have drawn the attention of researchers, since it could affect the levels of hormone circulation (growth hormone, melatonin, cortisol, prolactin, glucose, and insulin) which are major factors involved in numerous disease progression, including BC (Lu et al., 2017).

*Dietary Habits***:** The Nurses' Health Study and study conducted in NIH-AARP Diet and Health and also several other studies had observed heavy intake of animal fats, red meat, smoked meat and lower intake of fruits, vegetables and total dietary fiber to have significant association with BC risk (Haraldsdottir et al., 2018; Kim et al., 2017). Probable mechanisms for the increased risk include oxidative damage from bioavailable heme-iron, consumption of exogenous growth-promoting hormones through animal food production, exposure to mutagenic xenobiotic compounds such as N-Nitroso compounds (NOCs), polycyclic aromatic hydrocarbons (PAHs) and heterocyclic amines (HCAs) (Kim et al., 2016).

*Tobacco and alcohol intake:* Among the women of northeast India, BC is the most prevalent cancer and a significant association was found with chewing habits of betel quid and tobacco (Malvia et al., 2017). Smoking has a modest association with BC, but the risk of causing the disease is increased significantly. Smoking related BC risk is particularly high among those who started it from early adolescent or peri-menarcheal ages, and women with a family history of the BC (Jones et al., 2017). Extensive epidemiologic studies have linked moderate consumption of alcohol to increased risk of BC by approximately 30-50%. The risk of BC increases with the increase in the amount of alcohol consumption. Alcohol leads to changes in breast

density by altering the levels of estrogen thereby influencing BC risk (McDonald et al., 2013).

**C. Family history of BC**: The most well-established causal factor for the development of BC is family history. There is a two- fold increase BC risk for women having affected first- degree relatives. Familial relative risk (RR) increases with the age of diagnosis (<50 years) of first- degree relatives affected and also the number of relatives affected (Mavaddat et al., 2010; Brewer et al., 2017). Inheritance of mutation in high penetrance gene (*BRCA1* and *BRCA2*) is the main causal factor for family history of BC, but they account only a small fraction (10–20%) of all familial BC cases suggesting the contribution of other high, moderate and low penetrance genes that may take part in susceptibility to BC (Jian et al., 2017).

***2. Association of Histological tumor grade with other clinical and epidemiological features*:**

BC tumor expresses variation in molecular and morphological features, nature of action and response to therapy. At present, clinical management of BC depends on the information of several prognostic and predictive factors that help support patient in making decision and physicians to provide suitable and efficient treatment. Histological tumor grade is one the most established prognostic factors for BC since they could impart knowledge on morphological and biological characteristics of tumor (Rakha et al., 2010). Other equally important pathological and clinical features such as tumor type, lymph node and hormone receptor status

and Her2/Neu oncoproteins have been reported. They are known to have association with tumor grade and information on their collaborative effect could serves as an important prognostic factors (Atif et al., 2018).

Assessment of histological tumor grade for tubular differentiation composite, nuclear characteristics and mitotic activity are the most important parameters of BC evaluation. Particularly in early-stage BCs where no or less number of axillary lymph nodes are involved, the tumor grade serves as an important prognostic marker (Schwartz et al., 2014). Positive correlation has been reported between ER, PR and HER2/ neu status with tumor grade, and regardless of regional differences and ethnicity, an inverse relationship between ER, PR, HER2/neu and tumor grade was observed. With the increase in tumor grade, the expression of these receptors decreases (Thiygarajan et al., 2015; Siadati et al., 2015; Atif et al., 2018).

Lymph node status is another critical prognostic factor and help in determining the severity of the cancer, since cancer cells can metastasize to distant location (Pourzand et al., 2011; Siadati et al., 2015). Tumor grade was found to be significantly associated with the amount of axillary lymph nodes affected (Davis et al., 1986). With the increase in the amount of positive axillary lymph nodes, survival rate decreases and relapse rate increases (Siadati et al., 2015). Approximately 30-50% of BC cases diagnosed metastasize to the sentinel lymph node indicating the advanced cancer status (Pourzand et al., 2011).

The clinical and pathological features of breast tumor may differ between BC with family history and sporadic BC due to the inheritance of susceptibility gene

mutations. The inconsistent observation from different studies on familial BC characteristics remains controversial. Several research groups have revealed clinical features that are specific to familial cases. In fact, some studies had observed early age of disease onset, bilateral BC, advanced cancer stage, positive lymph node and lack of hormone receptors expression in familial cases, while others have not found any significant differences (Tazzite et al., 2013).

Even though several environmental factors such as lifestyle and dietary habits has been reported to contribute in BC progression, currently there is not much available data or literature focusing on the association between histological tumor grade and environmental factors. This study was carried out to find out whether the clinical and pathological features and the reproductive and environmental risk factors have association with histological tumor grade that might help in prediction and prevention from the related risk factors (Zodinpuii et al., 2019).

***3. Genetic Alteration in BC***:

With the evolving sequencing technologies, number of BC susceptibility genes has been reported other than *BRCA1* and *BRCA2* (Apostolou et al., 2013). These gene variants were broadly divided into two groups: first is the mutation in the proto-oncogenes (mostly gain-of-function that stimulate cell growth and proliferation) and second, is mutation occurring in the tumor suppressor genes (causing loss-of-function of the gene product). Loss of function results in uncontrollable growth of cell, incapability to repair DNA after damage and absence of cell cycle check points (Sheikh et al., 2015). Depending on their RR associated

with BC, these genes are classified into high, moderate and low penetrance genes (Muhammad et al., 2018).

High-penetrance genes are less common in the population but associated with higher risk and the carriers have a 5 to >20 RR. Moderate penetrance genes has a moderate association with BC risk and confer RR of 1.5 to 5 (Mavaddat et al., 2010; Apostolou et al., 2013). The involvement of the moderate penetrance genes to familial RR is < 3%, and the frequency of these gene variants are relatively low (Rahman et al., 2007). Commonly found in the population are low penetrance genes, but only has a small association with increase BC risk (RR <1.5) (Mavaddat et al., 2010; Apostolou et al., 2013). Chromosomal location of high penetrance genes, there associated syndrome and other related cancer risk is given in Table 1.

| Gene | Location | Syndrome | Other related cancers |
|---|---|---|---|
| BRCA1 | 17q21 | Hereditary Breast / Ovarian Cancer Syndrome | Prostate, Ovarian and fallopian tube, Melanoma, Biliary and Pancreas cancer |
| BRCA2 | 13q12-13 | | |
| TP53 | 17p13.1 | Li– Fraumeni Syndrome | Sarcomas, lung cancer, leukemia, adrenocortical carcinomas and brain cancer |
| PTEN | 10q23 | Cowden Syndrome PTEN Hamartoma Tumor Syndrome | GU tumor (renal cell carcinoma), Endometrail cancer and thyroid cancer |

| | | | |
|---|---|---|---|
| STK11 | 19p13.3 | Peutz-Jeghers Syndrome | Small bowel, colon, esophagus, stomach, pancreas, ovary, uterus, cervix, lung, and testis |
| CDH1 | 16q22.1 | Hereditary diffuse gastric cancer Syndrome | Gastric cancer, Colorectal cancer, diffuse subtype |

**Table 1: High penetrance BC genes and their associated cancers and syndromes.**

*BRCA1* and *BRCA2* are the two most crucial BC predisposing genes that have a major contribution in the development of inherited BC. Germline mutation carriers in these two high penetrant genes increases the lifetime risk of developing breast as well as ovarian cancer by 80% and are also known to be associated with the disease onset at an early age (D'Argenio et al., 2015). Mutations in these two genes are estimated to account only 15% of hereditary BCs cases (Shiovitz et al., 2015). *BRCA1/2* screening using direct Sanger sequencing for genetic testing becomes the most crucial component of clinical examination among female having a family history of breast and ovarian cancer. Next- generation sequencing (NGS) with its increased speed, efficiency in DNA test and reduced costs in comparison with Sanger sequencing had revolutionized genetic testing not only for a large size genes like *BRCA1/2*, but also for other multiple gene at a time (Walsh et al., 2010; Crawford et al., 2017).

After a landmark discovery of *BRCA1/2*, other genes that are linked to BC susceptibility have been identified and are often screened for mutations (Li et al.,

2018). Several studies performing a comprehensive multi-gene panel testing for BC has unraveled the remaining responsible genes other than *BRCA1/2*. A large number of gene mutations (~36- 61%) have been observed in BC patients including those among the high (*PTEN, CDH1, TP53, STK11 and NF1*), moderate (*NBN* and related genes, *CHEK2, PALB2, ATM*) and low penetrance gene (*PMS2, MEN1, MLH1, MSH6, PPM1D, MSH2* and so on) indicating that multiple gene panel testing serves as a more efficient tool for genetic cancer risk assessment for different sub-types of BC (Singh et al., 2018; Li et al., 2018). Moderate penetrance genes and their functions are listed in Table 2. Predisposition of multiple BC panel genes testing have been carried out extensively in some population such as Caucasian, African Americans, Slavs and Ashkenazi Jewish, however, data for other countries is still relatively low (Li et al., 2018).

| **Gene** | **Location** | **Function** |
|---|---|---|
| ATM | Xq22- Xq23 | Scrutinizing and repair dsDNA, phosphorylates multiple proteins including BRCA1, BRCA2, TP53 and CHEK2 |
| CHEK2 | 22q12.1 | Take part in cell cycle regulation at G2. React to DNA damage by Phosphorylation. Activated CHEK2 act to stabilize p53 and cooperates with BRCA1 |
| PALB2 | 16p12.2 | It promotes localization and stability of BRCA2 to facilitate BRCA2-mediated DNA repair. |

| | | |
|---|---|---|
| BRIP1 | 17q23.2 | It encodes a helicase which interacts with BRCA1 C-Terminus (BRCT) domain of BRCA1 and functions in checkpoint control |

**Table 2. The location and functions of Moderate penetrance BC genes.**

*TP53* mutation leads to Li– Fraumeni syndrome (LFS) and there is an increase of more than 90% lifetime cancer risk for *TP53* mutation carrier (Nichols et al., 2001; Gage et al., 2012; Apostolou et al., 2013). Among females *TP53* mutation carrier, BC incidence is the highest of all cancer, and ~5% of these BC cases are diagnosed at age less than 30 years (Apostolou et al., 2013; Mahdavi et al., 2019). *PTEN* is observed to be the second most altered gene in cancers. (Apostolou et al., 2013; Mahdavi et al., 2019). Various structural mutations were observed in highly aggressive type of BC (basal-like BC). Tumor in this type of BC, gross mutations of *PTEN* are greatly correlated with *BRCA1* mutations suggesting that *BRCA1* malfunction can directly target *PTEN* and its pathways, which may further describe why in *BRCA1* deficiency, basal-like BC often develop (Yin et al., 2008).

Mutations in *STK11* gene were found in 30-70% of Peutz-Jeghers syndrome (Schumacher et al., 2005). Peutz-Jeghers syndrome is strongly associated with various cancer types and carriers of *STK11* gene mutation confer up to 85% risk for developing any type of cancer. Female *STK11* gene mutations carrier are linked to an increased risk of about 20% for BC and estrogen-receptor positive type of BC are often found with these gene mutation (Apostolou et al., 2013; Mahdavi et al., 2019;

Sheikh et al., 2015). Hereditary diffuse gastric cancer syndrome is caused by mutation in *CDH1* gene (E-cadherin). Mutations in this gene also leads to an increase risk (~67% in male and 83% in female) of developing gastric cancer. (Keller et al., 1999; Pharoah et al., 2001). There is a 40%–54% lifetime risk of developing lobular BC for carriers of the *CDH1* gene mutations (Apostolou et al., 2013). Majority of mutations found in infiltrating lobular carcinoma are either frameshift or nonsense that result in the production of a dysfunctional E-cadherin molecule with reduced adhesion activity. Loss of heterozygosity (LOH) mutation in *CDH1* gene is also often found in lobular BCs. However, by far, there have been no reports of *CDH1* gene mutation in ductal carcinomas of the breast (Sheikh et al., 2015).

***Mitochondrial gene alterations in BC***

Mitochondria play an important role in BC pathogenesis. They produce energy through aerobic respiration, making mitochondria the main source and target of intracellular reactive oxygen species (ROS) (Thyagarajan et al., 2013). Mutation rates are much higher as compared to nuclear DNA most likely because of less efficiency in damage DNA repair mechanism (Parrella et al., 2001). Since the first report observed by Bianchi et al. (1995), a large number and variety of mitochondrial DNA (mtDNA) alterations was found in breast tumor tissue (Parr et al., 2006). Multiple studies have also found associations between cancer development and somatic mtDNA mutations, also hereditary mtDNA polymorphism have been found to contribute in cancer progression (Czarnecka et al., 2010). From the recent years, mtDNA have been screened for BC specific mutations simultaneously with nuclear genome. Mutations observed at certain positions (204, 207 and 16293) of mtDNA

that are obtained from nipple aspirate fluid have been an indicative for BC and mutations in mtDNA D-loop have been proposed as prognostic marker for BC independently. Studies have also shown that carriers of A10298G polymorphism are at an elevated risk for developing BC (Czarnecka et al., 2010). In studies performed by Zhu et al. (2005), the frequency of mutation is much higher in the mitochondrial D-loop region than in the other loci in 93% of the tested BC samples. Even though there is a wide variation in the observed frequencies, region of mtDNA characterized and the nature of mutations associated with breast tumor tissue, it is obvious that mutations in mtDNA serve as informative biomarkers for BC detection. A study on the mutations in the mitochondrial and nuclear DNA related with BC in the Mizo population was investigated to understand the mutations and their consequences for early diagnosis, proper treatment and counseling to the patients.

Despite Mizoram, being a state with high incidence of BC and lifestyle of high tobacco consumption and unique dietary habits, data on the epidemiological and genetic risk factors is limited. To date, there have been no reports that focus on the genetics or environmental risk factors towards BC for the Mizo population. The present study on Characterization of clinically significant mutations associated with Breast Cancer in Mizo population has been proposed with the aim of evaluating the demographic risk factors and assess the prevalence of mutation in susceptibility genes associated with BC in Mizo population, Mizoram. The findings of this study may help us to better understand the causal risk factors and the prevalence of certain gene mutations that might further help in prevention, early diagnosis, and provide best treatment option to the patients.

# Chapter II

# Aims and Objectives

## *Aims and Objectives*

The following objectives are set forth to carry out the proposed work:

- Study of mutations in candidate genes and their association with BC in Mizo Population.
- To determine the potential confounding factors associated with BC and their pathogenicity analysis using in- silico methods.

# Chapter III

# Materials and Methods

## *Materials and Methods*

### Sample collection

The aim and objectives of the study were first explained to the patient participants and with their consent, personal interview was performed using a bilingual structured questionnaire (English and Mizo local language). Information on reproductive history, environmental factors, tobacco and alcohol history, and family history with regard to breast, other cancers and inheritable diseases were collected. Patient clinical and pathological reports such as histological cancer type, grade of tumors, status of estrogen, progesterone, HER2 receptor and lymph node were acquired from their hospital documents using a standard proforma. Only local ethnic patient confirmed with BC of all ages from different hospitals of Mizoram, with and without family history of BC were included in the study. However, male BC patients were excluded from the study.

426 Breast Cancer (BC) patients from all over Mizoram and those registered in Mizoram State Cancer Institute (MSCI) within the year 2013 – 2017 and 810 healthy individuals were included in this study. 2 ml of blood was drawn from the participants (patients and healthy controls) by a trained technician and was stored in EDTA vials and kept in -20°C for further processing. Subject inclusion criteria are confirmed BC cases, with and without a history of BC in their family members, and BC patient of all ages from all over Mizoram, both registered as well as not registered in MSCI and belonging to Mizo tribe.

***1. Epidemiologic studies to determine the potential confounding factors associated with BC***

This study included a total of 1236 samples (426 BC cases and 810 healthy controls). The factors considered for the epidemiological analysis under reproductive history are age at marriage, marital status, age at menarche, parity, age at menopause, age at first delivery, number of children, breast feeding, duration of lactation, use of oral contraceptive pills and abortion. Several lifestyle habits including tobacco and alcohol intake and dietary habits were considered in the environmental factors. Also the familial history of BC, other cancer and inheritable diseases were included.

*Statistical analysis for epidemiologic studies*

Frequency distribution of different age group was calculated among the samples (both cases and controls). Univariate logistic regression analyses for Chi-square tests was performed to assess the independent risk factors association with BC risk. Significant independent factors ($p$ value $< 0.05$) obtained from the univariate analysis were further considered in the multivariate analysis using Cox regression (introducing all variables and terms of interactions) (Kotsopoulos et al. 2012). Receiver operating charcateristic (ROC) curve was plotted for those factors that are found to be significant in the multivariate analysis, dividing them in groups for larger and smaller test results to find the sensitivity and specificity of the factors to estimate the potential risk score (Kumar et al., 2011). IBM Statistical Package for Social

Sciences (SPSS), software version 22.0 for windows was employed to carry out all the statistical analyses.

*Formaldehyde estimation*

Formaldehyde estimation was carried out on fish samples obtained from four different sources such as Mizoram, Silchar, Burma and Andhra Pradesh. 60 ml of 6% w/w Tri- chloro acetic acid was added to 30 grams of finely chopped fish. The mixture was filtered using Whatman No.1 filter paper. 5ml of filtrate was collected (pH adjusted to 7 with NaOh or HCL) which was stored in deep freezer for 30 minutes to 1 hour. 2 ml of Nash reagent was added to the filtrate followed by incubation in water bath at $60^{o}$ C for 30 minutes. Absorbance was measured at 415 nm. Normal calibration curve for standard formaldehyde solution was prepared by making a serial dilution of 10 ppm stock solution of formaldehyde giving, 0.005, 0.1, 0.5, 1, 5 and 10 ppm concentration of standard solution. 2 ml of Nash's reagent was added to each dilution flask, which was incubated in water- bath at $60^{o}$ C for 30 minutes. Volume was maintained to 100 ml by adding distilled water. Absorbance was measured at 415 nm (Jaman et al., 2015; Sultana et al., 2018).

***2. Association of Histological tumor grade with other clinical and epidemiological features***

In this study, 103 BC cases with complete information on clinical and histological reports and the epidemiological data were included in the analysis. Clinical and histopathological records include tumor characteristics such as the

tumor side, grade, type, size, site, expression of hormone receptors (estrogen and progesterone), Her2/neu oncoprotein, and lymph node invasion. The histopathological grading of the BC tumor was done based on the Nottingham modification of the Bloom Richardson grading system accepted by the World Health Organization. Age range 41-50 years was observed to be having the highest incidence of BC, so, in order to perceive the best possible association between tumor grades with the factors considered, factors that do not show significant results in the first analysis were further analyzed by limiting the age group to 40-55 years which includes 51 cases out of the 103 cases.

IBM Statistical Package for Social Sciences (SPSS), software version 22.0 for Windows was utilized for all the statistical analyses. The frequency distribution was calculated for age at diagnosis and tumor characteristics. Statistical analysis for all the variables was done using chi square test. Since the analysis includes multiple subgroups, 5% level of significance ($p< 0.05$) was maintained (Kuzhan et al., 2013; Pourzand et al., 2011).

### *3. Study of mutations in candidate genes and their association with BC in Mizo Population*

*Genomic DNA Isolation*

QIAamp DNA mini kit (250), catalogue no. 51306 was used to extract genomic DNA from 200 µl of blood following the manufacturer's protocol. The extracted DNA samples were stored in -20$^{o}$C for further analysis.

### *3.1. Study of genetic alteration using Direct Sanger Sequencing*

50 cases and 50 healthy controls were included for this study and the candidate genes such as *BRCA1, TP53, PTEN, CDH1, CHEK2, XRCC2* were considered. The primers were designed to encompass the gene mutations reported for BC in the Human Gene Mutation Database (HGMD), except for the BRCA1 gene where all the 24 exons were sequenced.

The amplification was performed using specific primers designed for the specified region of interest (Table 3). Polymerase chain reaction (PCR) was performed using 25 µl total reaction volumes, each containing template DNA (100ng/µl), forward and reverse primers (10 pM/µl), 10X PCR buffer (2.5 µL), dNTPs (10mM), Taq polymerase (1U) and milli-q $H_2O$ (to make up the volume). PCR conditions used were as follows, initial denaturation at 95°C for 5 minutes, denaturation at 94°C for 30 seconds, annealing for 50-63°C for 35 seconds, extension at 72°C for 35 seconds and final extension at 72°C for 3 minutes. The PCR amplified products were loaded in 1.2% agarose gel and the success of amplification was checked using Gel documentation system (Bio-Rad, USA).

| **Gene** | **Exon** | **Primers (5'-3')** | **Product size (bp)** |
|---|---|---|---|
| BRCA1 | 2 | GAAGTTGATCATTTTATAAACCTTT<br>TGTCTTTTCTTCCCTAGTATGT | 258 |
| | 3 | TCCTGACACAGCAGACATTTA<br>TTGGATTTTCGTTCTCACTTA | 338 |
| | 5 | GTTGTGAGATTATCTTTTCATGGC<br>CTTCCAACCTAGCATCATTACCA | 208 |
| | 6 | CTTATTTTAGTGTCCTTAAAAGG<br>TTTCATGGACAGCACTTGAGTG | 206 |
| | 7 | CACAACAAAGAGCATACATAGG<br>AGAAGAAGAAGAAAACAAATGG | 269 |
| | 8 | TGTTAGCTGACTGATGATGGT | 267 |

| | | | |
|---|---|---|---|
| | | ATCCAGCAATTATTATTAAATAC | |
| | 9 | CCACAGTAGATGCTCAGTAAATA<br>TAGGAAAATACCAGCTTCATAGA | 211 |
| | 10 | TGGTCAGCTTTCTGTAATCG<br>GTATCTACCCACTCTCTTCTTCAG | 242 |
| | 11AB | TAGCCAGTTGGTTGATTTCC<br>CTCACACAGGGGATCAGCATTC | 477 |
| | 11C | CAACATAACAGATGGGCTGGAAG<br>ACGTCCAATACATCAGCTACTTTGG | 350 |
| | 11EF | GGTTCTGATGACTCACATGATGGG<br>TCATCACTTGACCATTCTGCTCC | 460 |
| | 11G | GAGCCACAGATAATACAAGAGCGTC<br>GCAGATTCTTTTTCGAGTGATTCTATTGG<br>G | 272 |
| | 11H | ATCAGGGAACTAACCAAACGGAG<br>CGCATGAATATGCCTGGTAGAAG | 269 |
| | 11J | CTAAAAAGAATAGGCTGAGGAGGAAGT<br>CAGCTCTGGGAAAGTATCGCTG | 284 |
| | 11K | GCAACTGGAGCCAAGAAGAGTAAC<br>TCTGTGTCATTTCTATTATCTTTGGA | 458 |
| | 11N | GCACTCTAGGGAAGGCAAAAACAG<br>CATTCCTCTTCTGCATTTCCTGG | 280 |
| | 11P | GCCAGTCATTTGCTCCGTTTTC<br>CGTTGCCTCTGAACTGAGATGATAG | 288 |
| | 11Q | TGCAGGCTTTCCTGTGGTTG<br>GGCTAATTGTGCTCACTGTACTTGG | 305 |
| | 11S | TCAATGTCACCTGAAAGAGAAATGG<br>CAGGATGCTTACAATTACTTCCAGG | 301 |
| | 11TU | TTGAATGCTATGCTTAGATTAGGGG<br>TTCTGAGGACTCTAATTTCTTGG | 402 |
| | 11V | GAGTCCTAGCCCTTTCACCCATAC<br>GTGATGTTCCTGAGATGCCTTTG | 289 |
| | 11W<br>X | CGTTGCTACCGAGTGTCTGTCTAAG<br>GTGCTCCCAAAAGCATAAA | 438 |
| | 12 | GTCCTGCCAATGAGAAGAAA<br>TGTCAGCAAACCTAAGAATGT | 265 |
| | 13 | AATGGAAAGCTTCTCAAAGTA<br>ATGTTGGAGCTAGGTCCTTAC | 320 |
| | 14 | CTAACCTGAATTATCACTATCA | 312 |

| | | | |
|---|---|---|---|
| | | GTGTATAAATGCCTGTATGCA | |
| | 15 | TGGCTGCCCAGGAAGTATG<br>AACCAGAATATCTTTATGTAGGA | 338 |
| | 16 | AATTCTTAACAGAGACCAGAAC<br>AAAACTCTTTCCAGAATGTTGT | 449 |
| | 17 | GTGTAGAACGTGCAGGATTG<br>TCGCCTCATGTGGTTTTA | 263 |
| | 18 | GGCTCTTTAGCTTCTTAGGAC<br>GAGACCATTTTCCCAGCATC | 351 |
| | 19 | CTGTCATTCTTCCTGTGCTC<br>CATTGTTAAGGAAAGTGGTGC | 249 |
| | 20 | ATATGACGTGTCTGCTCCACC<br>AATGAAGCGGCCCATCTC | 249 |
| | 21 | AAGCTCTTCCTTTTTGAAAGTC<br>GTAGAGAAATAGAATAGCCTCT | 298 |
| | 22 | TCCCATTGAGAGGTCTTGCT<br>GAGAAGACTTCTGAGGCTAC | 297 |
| | 23 | CAGAGCAAGACCCTGTCTC<br>ACTGTGCTACTCAAGCACCA | 255 |
| | 24 | ATGAATTGACACTAATCTCTGC<br>GTAGCCAGGACAGTAGAAGGA | 280 |
| TP53 | 5.6 | CGCTAGTGGGTTGCAGGA<br>CACTGACAACCACCCTTAAC | 550 |
| | 8.9 | GTTGGGAGTAGATGGAGCCT<br>GGCATTTTGAGTGTTAGACTG | 455 |
| PTEN | 4 | CTGTATTAGTGGCATCACAAGTTT<br>TGCACTTTAGTCTTCCTGACAA | 527 |
| | 5 | CCACAGTTGCACAATATCCTTT<br>CCAATAAATTCTCAGATCCAGGAAG | 301 |
| CDH1 | 9 | GACACATCTCTTTGCTCTGC<br>GGGACAAGGGTATGAACAGC | 269 |
| | 12 | GTCTGGTGGAAGGCAATGG<br>GAAGCATGGCAGTTGGAGC | 345 |
| CHEK2 | 10 | TGTCAACTGTTTGCTTGTCTTAATG<br>GCCAAGAAGAGAACAGCAAAC | 341 |
| | 12 | CTTGGACTGGCAGACTATGTT<br>ATGGTGGTGTGCATCTGTAG | 432 |
| XRCC2 | 2 | CAGCACCCAGCCTAAAGTTAT<br>AAGACAGAGGTCAAGGCATATT | 426 |

| | | | |
|---|---|---|---|
| | 3 | CAGCAGTCTACTCTGAGGAAATG<br>TGCAGTGAGCCATGATTGT | 457 |

**Table 3. Candidate Genes and their exons used in the present study.**

The details of the primer sequence and the expected product size are given.

*Purification, Sequencing and Sequence analysis*

Purification of the PCR amplified products to remove excess salts, primers and dNTP's was carried out using Qiagen purification kit and Exonuclease and shrimp Alkaline phosphatase (Exo- sap) following the standard manufacturer's protocol (Applied Biosystem, USA). Applied Biosystem, 3500 Genetic Analyzer was used to carry out sequencing from both the direction using forward and reverse primers for reading accuracy. Chromatographic files obtained were analyzed using FinchTV version 1.4.0 and was aligned using NCBI BLASTN (www.ncbi.nlm.nih.gov/blast). Ensembl (https://asia.ensembl.org/index.html) and HUGO Gene Nomenclature Committee (HGNC) (https://www.genenames.org) were used to check the amplification of the exonic regions. Mutation Taster (http://www.mutationtaster.org), BIC (Breast Cancer Information Core) (https://research.nhgri.nih.gov/bic/), ARUP BRCA (http://arup.utah.edu/database/BRCA/Home/BRCA1_landing.php), PolyPhen- 2 (http://genetics.bwh.harvard.edu/pph2/), Align GVGD (http://agvgd.hci.utah.edu) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases were used to analyze the pathogenicity and novelty of the observed polymorphisms. Structural prediction was done using HOPE software (https://www3.cmbi.umcn.nl/hope/).

### *3.2. Study of BC susceptibility gene BRCA1 and BRCA2 using Targeted re-sequencing*

This study included 10 early onset BC cases (having first or second degree relatives diagnosed with BC) and 5 healthy controls to identify, prioritize and rationalize clinically relevant mutation of *BRCA1* and *BRCA2* genes.

*Sequencing and Sequence processing*

TruSeq Custom Amplicon method was used to design the oligo probes specifically for the target regions of *BRCA1* and *BRCA2*, which uses Illumina Design Studio (Illumina, Inc., San Diego, CA, USA). A pair of oligo probes was designed for every 150 bp sequence of the focus region, in which the 5' and 3' ends of the sequence hybridize with the probe at one end and the other end of probe was complementary to the PCR primers. These oligo probes were then used for library construction, which contains the required nucleotide sequences. The target region comprises all the coding region of *BRCA1* and *BRCA2* genes, as well as 50 nucleotides upstream and downstream of the exon was considered to include the region of intron-exon. Sequencing was carried out using MiSeq Illumina sequencer (Illumina, Inc.)

The raw reads were trimmed against GRCh37/ hg19 (human Genome 19) using trimmomatic and was aligned using BWA-mem (http://bio-bwa.sourceforge.net/, version 0.7.10-r789) (Li, 2014). GATK (https://software.broadinstitute.org/gatk/, ver- sion 3.4–46) and picard tools (https:// broadinstitute.github.io/picard/, version 1.97) were used for variant calling as indicated in the workflow (Figure 1) (DePristo et al., 2011). Sequence analysis was

carried out using MiSeq software. Q-score of 30 was considered as an accepted threshold value that corresponds to 1:1,000 error rate. Mutations were categorized as SNPs (single nucleotide polymorphisms) and Indels (Insertions and deletions). Relevant mutations were query and analyzed using GeminiDB.

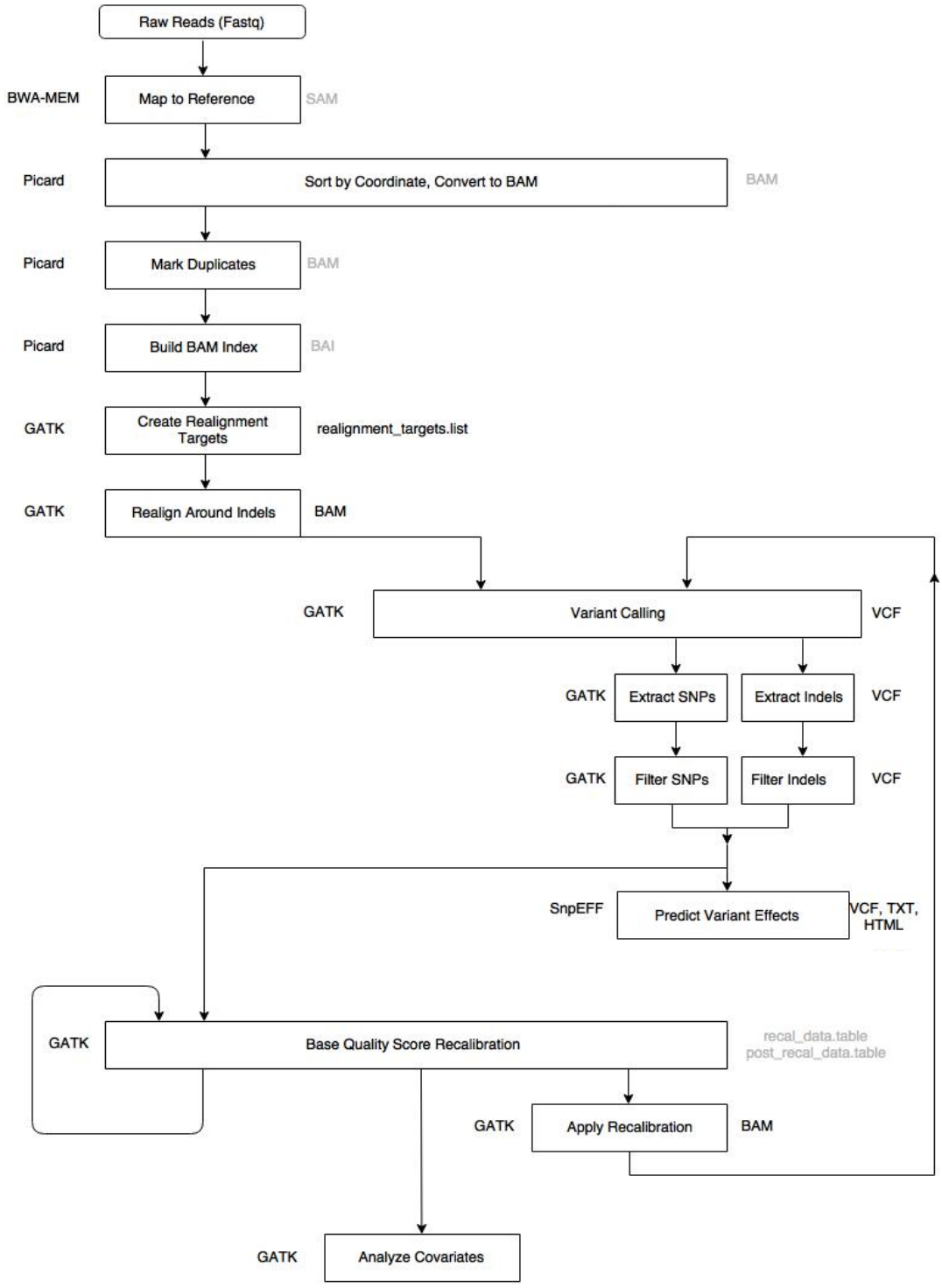



**Figure 1: Workflow used to identify the variants in the BC and healthy controls**

*Sequence Analysis*

Sequence variants obtained were analyzed using mutation prediction tools such as SIFT (http://sift.jcvi.org), Polyphen- 2 (http://genetics.bwh.harvard.edu/pph), LRT (http://www.genetics.wustl.edu/), Mutation taster (www.mutationtaster.org/) and databases such as BIC BRCA (Breast Cancer Information Core) (https://research.nhgri.nih.gov/bic/), ESP5400 (https://evs.gs.washington.edu/EVS/), Arup BRCA (https://arup.utah.edu/database/BRCA/), 1000g (https://www.internationalgenome.org) and dbSNP (https://www.ncbi.nlm.nih.gov/snp/).

PCR amplification of the desired region followed by Sanger sequencing was carried out to validate the presence of those variants having minor allele frequencies (MAF) < 0.05 and being classified as pathogenic based on the prediction tools. Confirmed variants were then analyzed for their pathogenicity using mutation taster, BIC BRCA, ARUP BRCA, BRCA Exchange (https://brcaexchange.org) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) databases were used to check the novelty of the variants.

### *3.3. Study on clinically significant mutations using Clinical Exome sequencing.*

Clinical exome sequencing covering 9,643 genes (28983612 bp) was performed for 12 samples (7 familial cases, 2 Non-familial cases and 3 healthy controls) using Illumina HiSeqX sequencing platform. The following bioinformatics

work flow was performed which contains major components categorized as primary, secondary and tertiary analysis:

A. Primary Analysis:- Raw FASTQ files were processed to remove adapters and generate QC metrics on raw and trimmed read data. The fastq-mcf command line tool was utilized to detect and remove the sequencing adapters, primers and poor quality nucleotides at the ends of reads (Andrews, 2010).

B. Secondary Analysis:- The trimmed FASTQ reads were aligned to a human genome reference sequence (GRCh37/hg19) in order to obtain variants. GATK is being used for secondary analysis, which is carried out as follows:

    i. Alignment of trimmed FASTQ reads using BWA-MEM (Li, 2014).

    ii. SAM to BAM file conversion using samtools (Li et al., 2010).

    iii. Removal of duplicates using Picard tools.

    iv. Indel realignment using GATK applications (DePristo et al., 2011).

    v. Variant calling was performed using GATK UnifiedGenotyper and Haplotypecaller (Shahi et al., 2019).

C. Tertiary analysis (Variant and gene annotation):- VariMAT (Variation and Mutation Annotation Toolkit) combines multiple clinical grade databases, variant class and pathogenicity prediction tools for annotating the variants and mutations that are relying on VEP was used for annotations (Gupta et al., 2017).

50 BC panel genes were retrieved from 11 companies such as Ambry Genetics, BreastHealth UK, University of Washington Centogene, Fulgent Diagnostics, GeneDx, Illumina, Invitae, Emory Genetics Laboratory, Myriad Genetics and CD Genomics. The panel genes are *ATM, ATK1, ABRAXAS, ATR, APC, AXINS,BARD1, BAP1, BLM, BMPR1A, BRCA2, BRIP1, CDK4, BRCA1, CDKN2A, CDH1, CHEK2, EPCAM, CTNNB1, FANCC, FANCM, FAM175A, GEN1, HOXB13, MEN1, MLH1, MRE11A, MUTYH, MSH2, MSH6, NBN, NF1, PALB2, PALLD, PIK3CA, PMS1, PMS2, RAD50, RAD51, RAD51C, PTEN, RAD51D, RECQL, RINT1, SMAD4, VHL, XRCC2, XRCC3, STK11, TP53* (Easton et al., 2015).

The VCF (Variant Call Format) file for each of the 12 samples were filtered by limiting only those variants belonging to the above-mentioned genes. Variants that are specific only to cases were selected and further analysis is done in a two- way approach as follows: First, case specific variants were filtered by strictly keeping only those of variant class such as missense, frameshift indel and intronic- SS-ACR/ DNR with a MAF< 0.05 based on databases such as ExAC (http://exac.broadinstitute.org), 1000G and dbSNP. Variant classification was done according to the American college of Medical Genetics and Genomics (ACMG) guidelines using Genetic Variant interpretation tools and literature search(https://www.medschool.umaryland.edugenetic_variant_interpretation_tools2.html/) (Li et al., 2018; Richards et al., 2015).

Second, Pathogenic variants were filtered based on the software prediction tools such as SIFT, Polyphen- 2, Provean (http://provean.jcvi.org/index.php),

LRT, FATHMM (http://fathmm.biocompute.org.uk), METASVM (https://omictools.com/meta-svm-tool), METALR (http://www.ensembl.info/tag/metalr/) and CADD (https://cadd.gs.washington.edu/score). Hope structural prediction was also performed for variants that are novel as well as predicted pathogenic by ACMG guidelines.

### *3.4. Study on mitochondrial gene alteration in BC using whole mitochondrial genome sequencing*

*Amplification of mtDNA and sequencing*

A total of 44 samples, comprising of 32 cases and 12 healthy controls were included in this study. The age range of the patients at diagnosis is 35 to 85 years. Complete mtDNA (16.6 kb) amplification from blood was carried out in two large overlapping amplicons of 9.3kb and 7.6 kb using two pairs of already published primers (Andrews S. 2010) in ABI 9700 thermal cycler (Thermo Scientific) using SequalPrepTM Long PCR Kit and dNTPs (Thermo Scientific). The PCR amplified products were visualized using 0.8% agarose gel electrophoresis and purified using QIAquick Gel Extraction Kit (Qiagen) according to the manufacturer protocol.

Sequencing libraries was prepared from 1 ng equimolar pools of gel purified amplicons using Nextera XT DNA library kit (Illumina). The quality and quantity of

the selected libraries were assessed using Picogreen dye in Qubit Fluorimeter (Invitrogen) and High Sensitivity Dna chip (Agilent) in 2100 Bioanalyzer (Agilent). Sequencing of the pooled libraries was performed in Illumina HiSeq – 2500 to generate 2 x 100 bp reads for an intended coverage of > 5000 X.

*Sequence Analysis*

Sequencing generated raw sequence data were converted to FASTQ files using CASAVA package (version1.8.2, Illumina). The mtDNA amplicon sequencing data generated in Illumina HiSeq-2500 produced FASTQ files, which was further analyzed using FASTQC (Andrews, 2010) for preliminary quality checking. The sequence data obtained was aligned against the mitochondrial reference sequence (rCRS- revised Cambridge Reference Sequence) using BWA- MEM (Li, 2013). Conversion of SAM file to BAM files was done using SAMtools (Li et al., 2010) and reads that are less than mapping quality of 40 were discarded. To ensure coverage of the entire mtDNA reference sequence, the quality of BAM files were checked using QualiMap (García-Alcade et al., 2012). Detection of germline variants in the mtDNA was done using VarScan2 (Homer). Variants base quality score (q) < 20 were not included. Variant calls not supported by at least 3% of reads or presented <10% of total reads in either direction were not included. Integrative Genome Viewer (Wang et at., 2010) and ANNOVAR (Kloss-Brandstätter et al., 2011) were used for visualization and annotation of variants, respectively.

Several mitochondrial genome databases such as mtDB (http://www.genpat.uu.se/mtDB/), MITOMAP (http://www.mitomap.org) and

mtSNP (http://www.mtsnp.tmig.or.jp/mtsnp/index_e.shtml) were used to check the annotated variants. MITOMASTER (https://www.mitomap.org/WebHome) was used to analyze the nucleotide variants relative to rCRS. Pathogenicity of the variants was analyzed using SIFT (https://sift.bii.a-star.edu.sg/) and PolyPhen-2 (genetics.bwh.harvard.edu/pph2/). Hydrophobic/ hydrophilic nature of mitochondrial proteins prediction was performed using SOSUI system (http://sosui.proteome.bio.tuat.ac.jp)21. Potential impacts of non-synonymous substitutions on proteins were predicted using Pmut (http://www.ics.uci.edu/~baldig/mutation.html). The ratio of Synonymous to Non-synonymous variants was calculated. Correlation analysis between the differences in frequency of variants (case and controls) and BC risk was calculated using Fisher exact test, where $p < \alpha$ (significance level: $\alpha = 0.05$) is considered to demonstrates a significant difference between the two groups. Simultaneously, the relative risk (RR) assessment was described by the odds ratio (OR) and 95 % confident interval (95 % CI). Using the p value of the chi-square test, the correlation between SNP and BC was analyzed so as to assessed according to the high risk evaluation index: OR = 1 (SNP is not associated with BC), OR > 1 (SNP is a high-risk factor for BC occurrence) and OR < 1 (SNP is a protective factor for BC). All statistical analyses were performed using the SPSS 20.0 version, (IBM corp, Armonk, NY). CIRCOS24 was plotted to represent the variant frequency within breast cancer cases as well as between the cases and healthy controls. Variant clustering on mitochondrial genome based on lifestyle habits was predicted using heatmap (https://biit.cs.ut.ee/clustvis/)

# Chapter IV

# Results

## *Results*

**To determine the potential confounding factors associated with BC and their pathogenicity analysis using in- silico methods.**

### *1. Epidemiologic studies to determine the potential confounding factors associated with BC*

The age range of the BC cases studied is 20- 91 years with a mean age of 49 years. Maximum numbers of cases are diagnosed at the age of 41- 50 years (Figure 2).

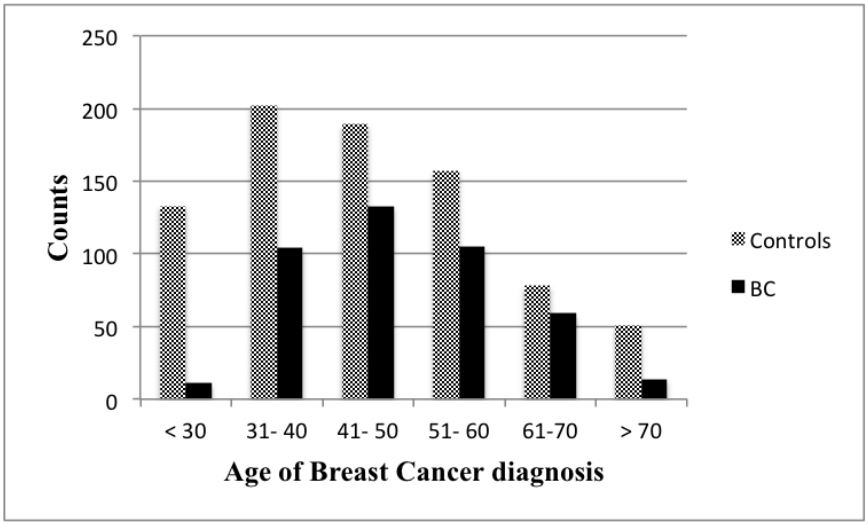



**Figure 2: Frequency distribution of study samples for epidemiologic studies**

From the univariate logistic regression analysis, reproductive factors such as marital status, age at menarche, age at menopause, parity, number of children, breast-feeding and duration of lactation were observed to be a potential risk factors for BC. Whereas, age at marriage, age at first delivery, birth control pills and abortion did not show any significant association (Table 4). Lifestyle factors like lack of exercise and lesser hour of sleep duration were also found to be the significant risk factors, whereas night shift work did not show any significance (Table 5). In dietary habits, regular or excess consumption of pork, fish, beef, saum, smoked food and oil, and less intake of vegetables and water showed significance indicating their positive contribution for BC progression. While chicken, fruits, and salt intake did not show any significance (Table 6). Consumption of tobacco related products such as sahdah, khaini and tuibur were found to play important role for the development of BC whereas, smoking cigarette and alcohol consumption did not show any significance (Table 7). Inheritable diseases and having first or second degree relatives with BC was found to be statistically significant for causing BC (Table 8).

| REPRODUCTIVE HISTORY | | | | |
|---|---|---|---|---|
| **Risk factors** | **Cases N (%)** | **Controls N (%)** | **OR (95% CI)** | **P value** |
| **Age at marriage (years)** | | | | |
| >=31 | 41 (10.8%) | 60 (8.9%) | Ref | |
| 21-30 | 95 (25.1%) | 157 (23.2%) | 1.29 (0.84- 1.97) | 0.241 |
| <=20 | 243 (64.1%) | 459 (67.9%) | 1.14 (0.84- 1.54) | 0.380 |
| Total | 379 (100 %) | 676 (100 %) | | |
| **Marital Status** | | | | |
| Yes | 378 (88.7%) | 674 (83.2%) | Ref | |
| No | 48 (11.3%) | 136 (16.8%) | 1.58 (1.11- 2.26) | <0.010 |
| Total | 426 (100%) | 810 (100%) | | |
| **Age at menarche (years)** | | | | |
| >15 | 132 (31.0%) | 208 (25.7%) | Ref | |
| <=15 | 294 (69.0%) | 602 (74.3%) | 1.29 (1.00- 1.68) | <0.047 |
| Total | 426 (100%) | 810 (100%) | | |
| **Age at menopause (years)** | | | | |
| <=50 | 230 (54.0%) | 245 (30.2%) | Ref | |
| >=51 | 196 (46.0%) | 565 (69.8%) | 2.70 (2.12- 3.44) | <0.000 |
| Total | 426 (100%) | 810 (100%) | | |
| **Parity** | | | | |
| Parous | 362 (85.0%) | 650 (80.2%) | Ref | |
| Nulliparous | 64 (15.0%) | 160 (19.8%) | 1.39 (1.01- 1.91) | <0.041 |
| Total | 426 (100%) | 810 (100%) | | |
| **Age at first delivery (years)** | | | | |
| <=20 | 113 (31.2%) | 204 (31.4%) | Ref | |
| 21-30 | 212 (58.6%) | 391 (60.2%) | 0.82 (0.51- 1.32) | 0.424 |
| >=31 | 37 (10.2%) | 55 (8.5%) | 0.80 (0.51- 1.26) | 0.346 |
| Total | 362 (100%) | 650 (100) | | |
| **No of Children** | | | | |
| 10-12 | 1 (0.3%) | 4 (0.6%) | Ref | |
| 7-9 | 15 (4.1%) | 24 (3.7%) | 0.49 (0.05-4.45) | 0.531 |
| 4-6 | 133 (36.7%) | 199 (30.7% | 1.23 (0.63- 2.41) | 0.529 |
| 1-3 | 213 (58.8%) | 422 (65.0%) | 1.32 (1.00-1.74) | <0.045 |
| Total | 362 (100%) | 649 (100%) | | |
| **Breast Feeding** | | | | |
| Yes | 357 (83.8%) | 638 (78.8%) | Ref | |
| No | 69 (16.2%) | 172 (21.2%) | 1.39 (1.02- 1.89) | <0.034 |
| Total | 426 (100%) | 810 (100%) | | |
| **Lactation Duration (Months)** | | | | |
| >=181 | 12 (3.4%) | 13 (2.0%) | Ref | |

| | | | | |
|---|---|---|---|---|
| 121-180 | 18 (5.0%) | 30 (4.7%) | 1.97 (0.88- 4.40) | 0.097 |
| 61-120 | 138 (38.7%) | 193 (30.2%) | 1.28 (0.69- 2.35) | 0.423 |
| 1-60 | 189 (52.9%) | 404 (63.1%) | 1.52 (1.15- 2.01) | <0.003 |
| Total | 357 (100%) | 640 (100%) | | |
| **Birth Control Pills** | | | | |
| No | 331 (77.7%) | 657 (81.1%) | Ref | |
| Yes | 95 (22.3%) | 153 (18.9%) | 0.81 (0.60- 1.08) | 0.155 |
| Total | 426 (100%) | 810 (100%) | | |
| **Abortion** | | | | |
| No | 319 (74.9%) | 608 (75.1%) | Ref | |
| 1 | 85 (20.0%) | 143 (17.7%) | 1.40 (0.84- 2.33) | 0.188 |
| >=2 | 22 (5.2%) | 59 (7.3%) | 1.59 (0.91- 2.78) | 0.102 |
| Total | 426 (100%) | 810 (100%) | | |

**Table 4: Reproductive History and their association with BC**

| **LIFESTYLE FACTORS** | | | | |
|---|---|---|---|---|
| **Risk Factors** | **Cases N (%)** | **Controls N (%)** | **OR (95% CI)** | **P value** |
| **Exercise** | | | | |
| Everyday | 123 (28.9%) | 220 (27.2%) | Ref | |
| 3-4 times | 111 (26.1%) | 171 (21.1%) | 1.22 (0.92- 1.61) | 0.162 |
| Never | 192 (45.1%) | 419 (51.7%) | 1.41 (1.05- 1.90) | <0.020 |
| Total | 426 (100%) | 810 (100%) | | |
| **Sleeping Hours** | | | | |
| >6 hrs | 264 (62.0%) | 579 (71.5%) | Ref | |
| <6 hrs | 162 (38.0%) | 231 (28.5%) | 0.65 (0.50- 0.83) | <0.001 |
| Total | 426 (100%) | 810 (100%) | | |
| **Night Duty** | | | | |
| No | 361 (84.7%) | 718 (88.6%) | Ref | |
| Yes | 65 (15.3%) | 92 (11.4%) | 0.71 (0.50- 1.00) | <0.051 |
| Total | 426 (100%) | 810 (100%) | | |

**Table 5: Lifestyle factors and their association with BC**

| DIETARY HABITS | | | | |
|---|---|---|---|---|
| **Risk Factors** | **Cases N (%)** | **Controls N (%)** | **OR (95% CI)** | **P value** |
| **Pork** | | | | |
| Never | 45 (10.6%) | 85 (10.5%) | Ref | |
| Thrice | 346 (81.2%) | 688 (84.9%) | 0.56 (0.31- 1.00) | 0.052 |
| Everyday | 35 (8.2%) | 37 (4.6%) | 0.53 (0.32- 0.85) | <0.010 |
| Total | 426 (100%) | 810 (100%) | | |
| **Fish** | | | | |
| Never | 22 (5.2%) | 59 (7.3%) | Ref | |
| Thrice | 277 (65.0%) | 582 (71.9%) | 0.49 (0.28- 0.85) | <0.011 |
| Everyday | 127 (29.8%) | 169 (20.9%) | 0.63 (0.48- 0.83) | <0.001 |
| Total | 426 (100%) | 810 (100%) | | |
| **Chicken** | | | | |
| Never | 16 (3.8%) | 43 (5.3%) | Ref | |
| Thrice | 279 (65.5%) | 544 (67.2%) | 0.63 (0.34- 1.16) | 0.144 |
| Everyday | 131 (30.8%) | 223 (27.5%) | 0.87 (0.67- 1.13) | 0.305 |
| Total | 426 (100%) | 810 (100%) | | |
| **Beef** | | | | |
| Never | 68 (16.0%) | 137 (16.9%) | Ref | |
| Thrice | 260 (61.0%) | 542 (66.9%) | 0.66 (0.44- 0.98) | <0.040 |
| Everyday | 98 (23.0%) | 131 (16.2%) | 0.64 (0.47- 0.86) | <0.004 |
| Total | 426 (100%) | 810 (100%) | | |
| **Fruits** | | | | |
| Regular | 135 (31.7%) | 181 (22.3%) | Ref | |
| Normal | 278 (65.3%) | 600 (74.1%) | 1.66 (0.83- 3.32) | 0.149 |
| Never | 13 (3.1%) | 29 (3.6%) | 1.03 (0.52- 2.01) | 0.923 |
| Total | 426 (100%) | 810 (100%) | | |
| **Vegetables** | | | | |
| Regular | 340 (79.8%) | 702 (86.7%) | Ref | |
| Occasional | 86 (20.2%) | 108 (13.3%) | 0.60 (0.44- 0.83) | <0.002 |

| | | | | |
|---|---|---|---|---|
| Total | 426 (100%) | 810 (100%) | | |
| **Sa-um** | | | | |
| Never | 88 (20.7%) | 135 (16.7%) | Ref | |
| Normal | 283 (66.4%) | 623 (76.9%) | 0.61 (0.38- 0.98) | <0.041 |
| Regular | 55 (12.9%) | 52 (6.4%) | 0.42 (0.28- 0.64) | <0.000 |
| Total | 426 (100%) | 810 (100%) | | |
| **Smoked Food** | | | | |
| Never | 43 (10.1%) | 121 (14.9%) | Ref | |
| Consumer | 383 (89.9% ) | 689 (85.1%) | 0.63 (0.44- 0.92) | <0.018 |
| Total | 426 (100%) | 810 (100%) | | |
| **Salt Intake** | | | | |
| Less | 142 (33.3%) | 293 (36.2%) | Ref | |
| Heavy | 284 (66.7%) | 517 (63.8%) | 0.88 (0.68- 1.13) | 0.321 |
| Total | 426 (100%) | 810 (100%) | | |
| **Water Intake** | | | | |
| >2 Ltrs | 49 (11.5%) | 72 (8.9%) | Ref | |
| 1-2 Ltrs | 125 (29.3%) | 313 (38.6%) | 1.14 (0.77- 1.70) | 0.494 |
| <1 Ltr | 252 (59.2%) | 425 (52.5%) | 0.67 (0.52- 0.87) | <0.003 |
| Total | 426 (100%) | 810 (100%) | | |
| **Oil Intake** | | | | |
| Less | 130 (30.5%) | 267 (33.0%) | Ref | |
| Normal | 146 (34.3%) | 320 (39.5%) | 0.72 (0.53- 0.97) | 0.032 |
| Heavy | 150 (35.2%) | 223 (27.5%) | 0.67 (0.51- 0.90) | <0.008 |
| Total | 426 (100%) | 810 (100%) | | |

**Table 6: Dietary habits and their association with BC**

| TOBACCO AND ALCOHOL HISTORY | | | | |
|---|---|---|---|---|
| **Risk factors** | **Cases N (%)** | **Controls N (%)** | **OR (95% CI)** | **P value** |
| **Sahdah consumption** | | | | |
| Never | 150 (35.2%) | 360 (44.4%) | Ref | |
| Consumer | 276 (64.8%) | 450 (55.6%) | 0.679 (0.533- 0.866) | <0.002 |
| Total | 426 (100%) | 810 (100%) | | |
| **Khaini consumption** | | | | |
| Never | 349 (81.9%) | 756 (93.3%) | Ref | |
| Consumer | 77 (18.1%) | 54 (6.7%) | 0.324 (0.224- 0.469) | <0.000 |
| Total | 426 (100%) | 810 (100%) | | |
| **Tuibur consumption** | | | | |
| Never | 297 (69.7%) | 632 (78.0%) | Ref | |
| Consumer | 129 (30.3%) | 178 (22.0%) | 0.648 (0.497- 0.845) | <0.001 |
| Total | 426 (100%) | 810 (100%) | | |
| **Cigarette consumption** | | | | |
| Non Smoker | 306 (71.8%) | 613 (75.7%) | Ref | |
| Smoker | 120 (28.2%) | 197 (24.3%) | 0.819 (0.629- 1.068) | 0.141 |
| Total | 426 (100%) | 810 (100%) | | |
| **Alcohol consumption** | | | | |
| Never | 419 (98.4%) | 786 (97.0%) | Ref | |
| Consumer | 7 (1.6%) | 24 (3.0%) | 1.828 (0.781- 4.277) | 0.164 |
| Total | 426 (100%) | 810 (100%) | | |

**Table 7: Tobacco and Alcohol consumption and their association with BC**

| FAMILY HISTORY IN RELATION TO CANCER | | | | |
|---|---|---|---|---|
| **Risk factors** | **Cases N (%)** | **Controls N (%)** | **OR (95% CI)** | **P value** |
| **1st and 2nd degree relative with breast cancer** | | | | |
| No | 333 (78.2%) | 724 (89.4%) | Ref | |
| yes | 93 (21.8%) | 86 (10.6%) | 0.425 (0.309- 0.586) | <0.000 |
| Total | 426 (100%) | 810 (100%) | | |
| **1st and 2nd degree relative with ovarian cancer** | | | | |
| No | 368 (86.4%) | 707 (87.3%) | Ref | |
| 1-2 | 58 (13.6%) | 103 (12.7%) | 0.924(0.654-1.306) | 0.655 |
| Total | 426 (100%) | 810 (100%) | | |
| **1st and 2nd degree relative with other cancer** | | | | |
| No | 149 (35.0%) | 350 (43.2%) | Ref | |
| 1-5 | 273 (64.1%) | 458 (56.5%) | 0.213(0.039-1.175) | 0.076 |
| 6-10 | 4 (0.9%) | 2 (0.2%) | 0.298(0.054-1.638) | 0.164 |
| Total | 426 (100%) | 810 (100%) | | |
| **Inheritable Disease** | | | | |
| No | 276 (64.8%) | 645 (79.6%) | Ref | |
| 1-4 | 150 (35.2%) | 165 (20.4%) | 0.471(0.362-0.612) | <0.000 |
| Total | 426 (100%) | 810 (100%) | | |

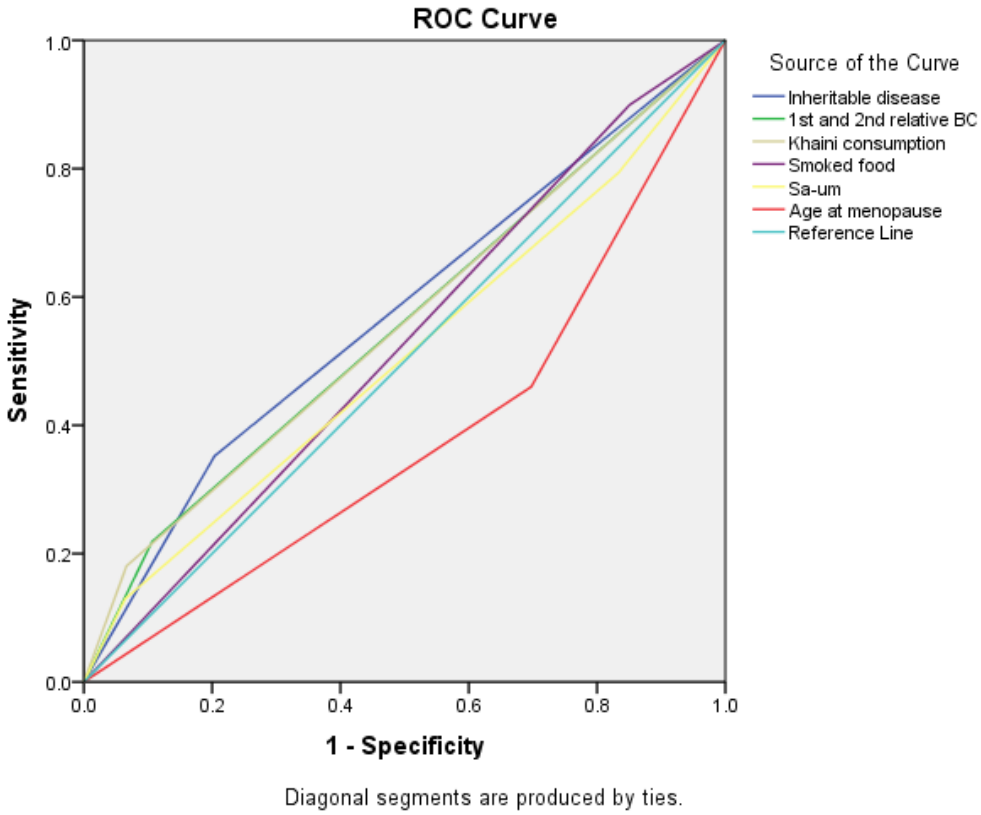
**Table 8: Family History in relation to hereditary diseases and cancer**

Multivariate analysis of potentialy important factors from univariate analysis showed that the risk of developing BC is found to be higher among those individuals with late age at menopause odds ratio (OR)= 0.356, 95% confidence interval (CI), 0.260- 0.488), shorter lactation duration (OR= 0.546, 95% CI, 0.395- 0.756), lack of physical exercise (OR= 0.506, 95% CI, 0.345- 0.742), lesser water intake (OR= 1.640, 95% CI, 1.178- 2.283) and also vegetables (OR= 1.989, 95% CI, 1.329- 2.977), high amount of sa-um consumption (OR= 1.930, 95% CI, 1.143- 3.260), smoked food (OR= 1.724, 95% CI, 1.070- 2779) and khaini (OR= 2.800, 95% CI, 1.726- 4.543), having 1$^{st}$ and 2$^{nd}$ degree relative with BC (OR= 2.458, 95% CI, 1.671- 3.618) and other inheritable diseases (OR= 2.693, 95% CI, 1.925- 3.769) were observed to have a significant association with BC risk (Table 9).

| Risk Factors | Odds Ratio (95% CI) | P value |
|---|---|---|
| Age at Menopause | 0.356 (0.260- 0.488) | < 0.000 |
| Lactation Duration | 0.546 (0.395- 0.756) | < 0.000 |
| Exercise | 0.506 (0.345- 0.742) | < 0.000 |
| Vegetables | 1.989 (1.329- 2.977) | < 0.001 |
| Sa-um | 1.930 (1.143- 3.260) | < 0.014 |
| Smoked Food | 1.724 (1.070- 2779) | < 0.025 |
| Water intake | 1.640 (1.178- 2.283) | < 0.003 |
| Khaini consumption | 2.800 (1.726- 4.543) | < 0.000 |
| 1$^{st}$ and 2$^{nd}$ degree BC relative | 2.458 (1.671- 3.618) | < 0.000 |
| Inheritable Diseases | 2.693 (1.925- 3.769) | < 0.000 |

**Table 9: Multivariate analysis for the significant risk factors**

The Receiver Operating Characteristics (ROC) curves plotted by selecting larger test results to indicate more positive test shows that the inheritable diseases (AUC: 0.574, 95% CI: 0.540- 0.608), having 1$^{st}$ and 2$^{nd}$ degree relatives with BC (AUC: 0.556, 95% CI: 0.522- 0.591), high consumption of khaini (tobacco) (AUC: 0.557, 95% CI: 0.522- 0.592), smoked food (AUC: 0.524, 95% CI: 0.491- 0.558) and sa-um (fermented pork fats) (AUC: 0.508, 95% CI: 0.473- 0.543) are potential significant risk factors for BC, while late age at menopause was found to act as a potential confounding factors among the Mizo BC patient (Figure 3).

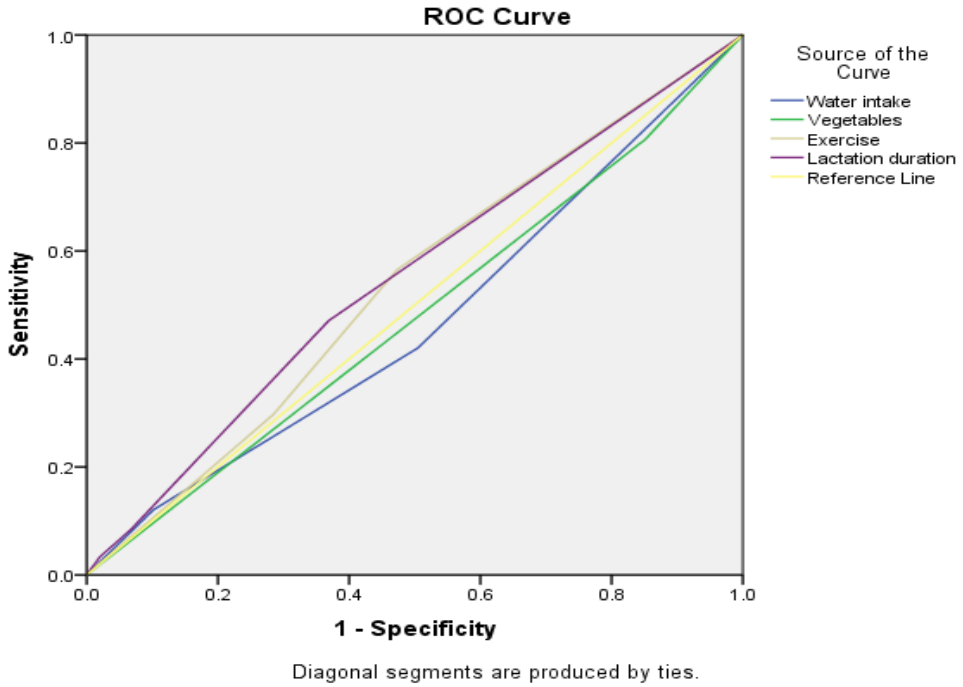



**Area Under the Curve**

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| Inheritable disease | .574 | .017 | .000 | .540 | .608 |
| 1st and 2nd relative BC | .556 | .018 | .001 | .522 | .591 |
| Khaini consumption | .557 | .018 | .001 | .522 | .592 |
| Smoked food | .524 | .017 | .161 | .491 | .558 |
| Sa-um | .508 | .018 | .628 | .473 | .543 |
| Age at menopause | .381 | .017 | .000 | .348 | .415 |

**Figure 3: ROC curve for larger values to indicate more positive test.**

By selecting smaller test results to indicate more positive test on ROC curve, we observed lack of physical exercises (AUC: 0.536, 95% CI: 0.499- 0.573) and shorter lactation duration (AUC: 0.551, 95% CI: 0.514- 0.588) to play a major role for BC development. Whereas, lesser intake of water and vegetables were found to act as a potential confounding factors (Figure 4).




**Area Under the Curve**

| Test Result Variable(s) | Area | Std. Error[a] | Asymptotic Sig.[b] | Asymptotic 95% Confidence Interval | |
|---|---|---|---|---|---|
| | | | | Lower Bound | Upper Bound |
| Water intake | .467 | .019 | .082 | .429 | .504 |
| Vegetables | .478 | .019 | .240 | .440 | .515 |
| Exercise | .536 | .019 | .059 | .499 | .573 |
| Lactation duration | .551 | .019 | .008 | .514 | .588 |

**Figure 4: ROC curve for smaller values to indicate more positive test**

From the formaldehyde estimation, fish coming from Andhra Pradesh was observed to have the highest concentration (3.17 µg/g), followed by Silchar (1.52 µg/g), Burma (0.28 µg/g) and Mizoram (0.17 µg/g).

***2. Association of Histological tumor grade with other clinical and epidemiological features***

This study includes samples with age range 23- 67 years with mean age 48 years. Infiltrating ductal carcinoma prevalence is found to be higher than infiltrating lobular carcinoma. Size of the tumor ranges from 1- 20.5 cm, with size < 5cm found to be most prevalent (86.40%). The independent expression of ER+, PR+ and HER2/neu+ were found to be almost similar with 54.36%, 49.51% and 49.51%, respectively. Expression of hormone receptor double subtype (ER+/PR+) with 46.60% is most common in grade I tumor whereas, ER-/PR- (42.71%) is more prevalent in grade II tumor. ER+/PR- (7.76%) or ER-/PR+ (2.91%) are found in 10% of the cases with tumor grade I and II and was not observed in tumor grade III. ER+/PR+/HER2- (27.18%) was found to be the most observed triple subtype, followed by ER-/PR-/HER2+ (25.24%), ER+/PR+/HER2+ (19.41%) and ER-/PR-/HER2- (17.47%). Distribution of clinico- pathological features of BC tumor is summarized in Table 10.

| Clinical and Histological features | Mean (min-max) | No. of Cases (%) |
|---|---|---|
| **Age at diagnosis (Years)** | 48 (23-67) | 103 |
| **Histological subtype** | | |
| Infiltrating Ductal Carcinoma | | 101 (98%) |
| Infiltrating Lobular Carcinoma | | 2 (1.9%) |
| **Pathological tumor size** | 3.44 (1-20.5) | |
| 0-5 | 2.671 (1-5) | 89 (86.40%) |
| 5-10 | 7.075 (5.4- 9.5) | 12 (11.65%) |
| >10 | 16.25 (12- 20.5) | 2 (1.94%) |
| **Lymph node invasion** | | |
| Positive | | 54 (52.42%) |
| Negative | | 49 (47.57%) |
| **Histological grade** | | |
| Grade I | | 43 (41.74%) |
| Grade II | | 48 (46.60%) |
| Grade III | | 12 (11.65%) |
| **Hormone receptor status** | | |
| ER +ve | | 56 (54.36%) |
| ER –ve | | 47 (45.63%) |
| PR +ve | | 51 (49.51%) |
| PR –ve | | 52 (50.48%) |
| Her2 +ve | | 51 (49.51%) |
| Her2 –ve | | 52 (50.48%) |
| ER+/PR + | | 48 (46.60%) |
| ER+/PR- | | 8 (7.76%) |
| ER-/PR - | | 44 (42.71%) |
| ER-/PR+ | | 3 (2.91%) |
| ER+/PR+/HER2+ | | 20 (19.41%) |
| ER+/PR+/HER2- | | 28 (27.18%) |
| ER+/PR-/HER2+ | | 3 (2.91%) |
| ER+/PR-/HER2- | | 5 (4.85%) |
| ER-/PR-/HER2+ | | 26 (25.24%) |
| ER-/PR-/HER2- | | 18 (17.47%) |
| ER-/PR+/HER2+ | | 2 (1.94%) |
| ER-/PR+/HER2- | | 1 (0.97%) |

**Table 10: Clinical and pathological features of BC tumor samples**

From the analysis of tumor grade and their association with lymph node invasion, tumor characteristics, ER, PR and HER2/neu oncoprotein- we have observed that with the increase in tumor grade, there is more of positive lymph node invasion. Significant association was found between the tumor grade and lymph node invasion ($p<0.021$), while no association was observed between tumor grade with the type, side, size and site of tumor. We also observed an association with ER ($p < 0.004$) and Her2/neu ($p< 0.014$) independently, expression of double subtype ER/PR ($p< 0.007$) and triple subtype ER/PR/HER2 ($p< 0.025$) with tumor grade. We also observed less expression of ER in higher grade of tumor (Table 11).

| **Factors** | **Grade I (n=43)** | **Grade II (n=48)** | **Grade III (n=12)** | **p-value** |
|---|---|---|---|---|
| **Lymph node invasion** | | | | |
| (+ve) | 17 (39.53%) | 27 (56.25%) | 10 (83.33%) | 0.021 |
| (-ve) | 26 (60.46%) | 21 (43.75%) | 2 (16.67%) | |
| **Tumor type** | | | | |
| IDC | 42 (97.67%) | 48 (100.00%) | 11 (91.67%) | 0.169 |
| ILC | 1 (2.32%) | 0 (0.00%) | 1 (8.33%) | |
| **Tumor side** | | | | |
| Left | 17 (39.53%) | 30 (62.50%) | 5 (41.67%) | 0.228 |
| Right | 25 (58.13%) | 17 (35.42%) | 7 (58.33%) | |
| Both | 1 (2.32%) | 1 (2.08%) | 0 (0.00%) | |
| **Tumor site** | | | | |
| All quadrant | 0 (0.00%) | 2 (4.17%) | 1 (8.33%) | 0.700 |
| Retro-areola | 4 (9.30%) | 5 (10.42%) | 1 (8.33%) | |
| LIQ/LOQ | 9 (20.93%) | 10 (20.83%) | 1 (8.33%) | |
| UIQ/UOQ | 30(69.76%) | 31(64.58%) | 9(75.00%) | |

| | | | | |
|---|---|---|---|---|
| **Tumor size (mm)** | | | | |
| 1-5 | 37 (86.04%) | 42 (87.50%) | 10 (83.33%) | 0.440 |
| 5-10 | 6 (13.95%) | 5 (10.42%) | 1 (8.33%) | |
| >10 | 0 (0.00%) | 1 (2.08%) | 1 (8.33%) | |
| **ER+/-** | | | | |
| +ve | 31 (72.09%) | 22 (45.83%) | 3 (25.00%) | 0.004 |
| -ve | 12 (27.91%) | 26 (54.17%) | 9 (75.00%) | |
| **PR+/-** | | | | |
| +ve | 24 (55.81%) | 24 (50.00%) | 3 (25.00%) | 0.168 |
| -ve | 19 (44.19%) | 24 (50.00%) | 9 (75.00%) | |
| **HER2+/-** | | | | |
| +ve | 14 (32.56%) | 30 (62.50%) | 7 (58.33%) | 0.014 |
| -ve | 29 (67.44%) | 18 (37.50%) | 5 (41.67%) | |
| **Er/Pr status** | | | | |
| ER+/PR+ | 24 (55.81%) | 21 (43.75%) | 3 (25.00%) | 0.007 |
| ER+/PR- | 7 (16.28%) | 1 (2.08%) | 0 (0.00%) | |
| ER-/PR+ | 0 (0.00%) | 3 (6.25%) | 0 (0.00%) | |
| ER-/PR- | 12 (27.91%) | 23 (47.92%) | 9 (75.00%) | |
| **Er/Pr/Her2 status** | | | | |
| Er+/Pr+/Her2+ | 6 (13.95%) | 12 (25.00%) | 2 (16.67%) | 0.025 |
| ER+/PR+/HER2- | 18 (41.86%) | 9 (18.75%) | 1 (8.33%) | |
| ER+/PR-/HER2+ | 3 (6.98%) | 0 (0.00%) | 0 (0.00%) | |
| ER+/PR-/HER2- | 4 (9.30%) | 1 (2.08%) | 0 (0.00%) | |
| ER-/PR+/HER2+ | 0 (0.00%) | 2 (4.17%) | 0 (0.00%) | |
| ER-/PR+/HER2- | 0 (0.00%) | 1 (2.08%) | 0 (0.00%) | |
| ER-/PR-/HER2+ | 5 (11.63%) | 16 (33.33%) | 5 (41.67%) | |
| ER-/PR-/HER2- | 7 (16.28%) | 7 (14.58%) | 4 (33.33%) | |

**Table 11: Correlation between tumor grade with lymph node invasion, tumor characteristics, hormone receptors and HER2/neu oncoprotein**

Analysis of different variables of reproductive factors showed that there is no association with tumor grade. 67.96% of BC cases included in the study were diagnosed at age >40 years. Our analysis showed that having at least one or more first degree relative's BC has a significant association with tumor grade ($p<0.003$). Other variables do not show any association with tumor grade (Table 12). From the analysis of the same factors, among 40-55 years at diagnosis, a consistent result was observed with having BC relative's (one or more first degree) ($p<0.029$) (Table 13).

| **Factors** | **Grade I (n=43)** | **Grade II (n=48)** | **Grade III (n=12)** | **p-value** |
|---|---|---|---|---|
| **Age range (mean)** | 49 | 46 | 49 | 0.482 |
| **Age at marriage (mean)** | 24 | 23 | 22 | 0.164 |
| **Age at menarche** | 15 | 15 | 15 | 0.286 |
| **Age at first delivery** | 24 | 25 | 23 | 0.225 |
| **No. of live birth** | 3 | 4 | 3 | 0.321 |
| **Lactation (months)** | 66 | 74 | 71 | 0.081 |
| **Age at diagnosis** | | | | |
| <40 | 12 (36.3%) | 16 (48.48%) | 5 (15.15%) | 0.642 |
| >40 | 31 (44.28%) | 32 (45.71%) | 7 (10.00%) | |
| **Marital status** | | | | |
| married | 39 (42.85%) | 40 (43.95%) | 12 (13.18%) | 0.225 |
| single | 4 (33.33%) | 8 (66.66%) | 0 (0.00%) | |
| **Menopausal** | | | | |
| Pre | 13 (37.14%) | 18 (51.42%) | 4 (11.42%) | 0.765 |
| Post | 30 (44.11%) | 30 (44.11%) | 8 (11.76%) | |
| **Parity** | | | | |
| parous | 38 (42.22%) | 40 (44.44%) | 12 (13.33%) | 0.211 |
| nulliparous | 5 (38.46%) | 8 (61.53%) | 0 (0.00) | |
| **Breast feeding** | | | | |

| | | | | |
|---|---|---|---|---|
| yes | 38 (42.22%) | 40 (44.44%) | 12 (13.33%) | 0.289 |
| no | 5 (38.46%) | 8 (61.53%) | 0 (0.00%) | |
| **Oral contraceptive pills** | | | | |
| Yes | 9 (34.61%) | 15 (57.69%) | 2 (7.69%) | 0.405 |
| no | 34 (44.15%) | 33 (42.85%) | 10 (12.98%) | |
| **Abortion** | | | | |
| never | 30 (38.46%) | 38 (48.71%) | 10 (12.82%) | 0.089 |
| once | 11 (57.89%) | 7 (36.84%) | 1 (5.26%) | |
| twice | 2(33.33%) | 3(50.00%) | 1(16.67%) | |
| **1st degree relative with breast cancer** | | | | |
| No | 41(95.35%) | 46(95.83%) | 10 (83.33%) | <0.003 |
| 1-2 | 2(4.65%) | 2(4.17%) | 0(0.00%) | |
| >3 | 0(0.00%) | 0(0.00%) | 2(16.67%) | |
| **2nd degree relative with breast cancer** | | | | |
| No | 38(88.37%) | 41(85.42%) | 10(83.33%) | 0.870 |
| 1 | 5(11.63%) | 7(14.58%) | 2(16.67%) | |
| **1st& 2nd degree relative with ovarian cancer** | | | | |
| No | 39(90.70%) | 40(83.33%) | 11(91.67%) | 0.511 |
| 1 and 2 | 4(9.30%) | 8(16.67%) | 1(8.33%) | |
| **1st& 2nd degree relative with other cancer** | | | | |
| No | 15(34.88%) | 20(41.67%) | 3(25.00%) | 0.219 |
| 1 and 2 | 25(58.14%) | 19(39.58%) | 8(66.67%) | |
| >3 | 3(6.98%) | 9(18.75%) | 1(8.33%) | |
| **Inheritable disease** | | | | |
| No | 30(69.77%) | 35(72.92%) | 8(66.67%) | 0.893 |
| 1-3 | 13(30.23%) | 13(27.08%) | 4(33.33%) | |

**Table 12: Correlation between tumor grade with reproductive and family history of BC**

| Factors | Grade I (n= 21) | Grade II (n= 25) | Grade III (n= 5) | p- value |
|---|---|---|---|---|
| **Age at marriage** | | | | |
| <25 | 13 (72.2%) | 14 (63.6%) | 5 (100.0%) | 0.267 |
| >25 | 5 (27.8%) | 8 (36.4%) | 0 (0.0%) | |
| **Age at menarche** | | | | |
| >16 | 7 (33.3%) | 13 (52.0%) | 2 (40.0%) | 0.440 |
| <15 | 14 (66.7%) | 12 (48.0%) | 3 (60.0%) | |
| **Age at 1st delivery** | | | | |
| <30 | 19 (90.5%) | 20 (80.0%) | 5 (100.0%) | 0.379 |
| >31 | 2 (9.5%) | 5 (20.0%) | 0 (0.0%) | |
| **Number of live birth** | | | | |
| >3 | 4 (23.5%) | 8 (38.1%) | 1 (20.0%) | 0.542 |
| <3 | 13 (76.5%) | 13 (61.9%) | 4 (80.0%) | |
| **Lactation duration (months)** | | | | |
| >72 months | 4 (23.5%) | 8 (38.1%) | 0 (0.0%) | 0.204 |
| <72 months | 13 (76.5%) | 13 (61.9%) | 5 (100.0%) | |
| **Age at diagnosis** | | | | |
| < 48 | 10 (47.6%) | 15 (60.0%) | 3 (60.0%) | 0.682 |
| > 49 | 11 (52.4%) | 10 (40.0%) | 2 (40.0%) | |
| **Marital status** | | | | |
| Married | 18 (85.7%) | 22 (88.0%) | 5 (100%) | 0.671 |
| Single | 3 (14.3%) | 3 (12.0%) | 0 (0.0%) | |
| **Age at menopause** | | | | |
| Pre | 5 (23.8%) | 8 (32.0%) | 2 (40.0%) | 0.716 |
| Post | 16 (76.2%) | 17 (68.0%) | 3 (60.0%) | |
| **Parity** | | | | |
| Parous | 17 (81.0%) | 21 (84.0%) | 5 (100%) | 0.574 |
| Nulliparous | 4 (19.0%) | 4 (16.0%) | 0 (0.0%) | |
| **Breast feeding** | | | | |
| Yes | 17 (81.0%) | 21 (84.0%) | 5 (100.0%) | 0.574 |

| | | | | |
|---|---|---|---|---|
| No | 4 (19.0%) | 4 (16.0%) | 0 (0.0%) | |
| **Oral contraceptive pills** | | | | |
| No | 18 (85.7%) | 20 (80.0%) | 4 (80.0%) | 0.870 |
| Yes | 3 (14.3%) | 5 (20.0%) | 1 (20.0%) | |
| **Abortion** | | | | |
| Never | 16 (76.2%) | 22 (88.0%) | 4 (80.0%) | 0.808 |
| Once | 4 (19.0%) | 2 (8.0%) | 1 (20.0%) | |
| Twice | 1 (4.8%) | 1 (4.0%) | 0 (0.0%) | |
| **1st degree relative with breast cancer** | | | | |
| No | 20 (95.2%) | 25 (100.0%) | 4 (80.0%) | <0.029 |
| 1-2 | 1 (4.8%) | 0 (0.0%) | 0 (0.0%) | |
| >3 | 0 (0.0%) | 0 (0.0%) | 1 (20.0%) | |
| **2nd degree relative with breast cancer** | | | | |
| No | 19 (90.5%) | 20 (80.0%) | 4 (80.0%) | 0.599 |
| 1 | 2 (9.5%) | 5 (20.0%) | 1 (20.0%) | |
| **1st and 2nd degree relative with ovarian cancer** | | | | |
| No | 19 (90.5%) | 20 (80.0%) | 4 (80.0%) | 0.599 |
| 1-2 | 2 (9.5%) | 5 (20.0%) | 1 (20.0%) | |
| **1st and 2nd degree relative with other cancer** | | | | |
| No | 8 (38.1%) | 7 (28.0%) | 2 (40.0%) | 0.566 |
| 1 and 2 | 11 (52.4%) | 12 (48.0%) | 3 (60.0%) | |
| >3 | 2 (9.5%) | 6 (24.0%) | 0 (0.0%) | |
| **Inheritable diseases** | | | | |
| No | 13 (61.9%) | 18 (72.0%) | 3 (60.0%) | 0.728 |
| 1-3 | 8 (38.1%) | 7 (28.0%) | 2 (40.0%) | |

**Table 13: Correlation between tumor grade with reproductive and family history of breast cancer among patients diagnosed at age between 40-55 years.**

The analysis of lifestyle habits and their impact on tumor grade was found to be non-significant for exercise, sleeping hours or night duty. From the analysis of dietary habits, despite significant differences observed in the intake amount and frequencies of the variables, there was no association observed between dietary habits and tumor grade. There was no significant association between tobacco related products and alcohol consumption with tumor grade (Table 14).

| Factors | Grade I (43) | Grade II (48) | Grade III (12) | p-value |
|---|---|---|---|---|
| **Pork** | | | | |
| Never | 2(4.65%) | 9(18.75%) | 1(8.33%) | 0.235 |
| Once | 23(53.49%) | 26(54.17%) | 5(41.67%) | |
| Thrice | 12(27.91%) | 11(22.92%) | 5(41.67%) | |
| everyday | 6(13.95%) | 2(4.17%) | 1(8.33%) | |
| **Fish** | | | | |
| Never | 0(0.00%) | 7(14.58%) | 1(8.33%) | 0.105 |
| Once | 29(67.44%) | 30(62.50%) | 7(58.33%) | |
| Thrice | 8(18.60%) | 8(16.67%) | 4(33.33%) | |
| everyday | 6(13.95%) | 3(6.25%) | 0(0.00%) | |
| **Chicken** | | | | |
| Never | 1(2.33%) | 4(8.33%) | 0(0.00%) | 0.378 |
| Once | 29(67.44%) | 33(68.75%) | 7(58.33%) | |
| Thrice | 7(16.28%) | 9(18.75%) | 4(33.33%) | |
| everyday | 6(13.95%) | 2(4.17%) | 1(8.33%) | |
| **Beef** | | | | |
| Never | 3(6.98%) | 9(18.75%) | 4(33.33%) | 0.230 |
| Once | 28(65.12%) | 29(60.42%) | 6(50.00%) | |
| Thrice | 6(13.95%) | 8(16.67%) | 1(8.33%) | |

| | | | | |
|---|---|---|---|---|
| everyday | 6(13.95%) | 2(4.17%) | 1(8.33%) | |
| **Fruits** | | | | |
| Regular | 15(34.88%) | 17(35.42%) | 4(33.33%) | 0.785 |
| Normal | 27(62.79%) | 27(56.25%) | 7(58.33%) | |
| Never | 1(2.33%) | 4(8.33%) | 1(8.33%) | |
| **Vegetables** | | | | |
| Regular | 36(83.72%) | 32(66.67%) | 8(66.67%) | 0.243 |
| Normal | 7(16.28%) | 12(25.00%) | 3(25.00%) | |
| Never | 0(0.00%) | 4(8.33%) | 1(8.33%) | |
| **Sa-um** | | | | |
| Never | 8(18.60%) | 12(25.00%) | 4(33.33%) | 0.796 |
| Normal | 27(62.79%) | 28(58.33%) | 7(58.33%) | |
| regular | 8(18.60%) | 8(16.67%) | 1(8.33%) | |
| **Smoked meat** | | | | |
| Never | 5(11.63%) | 7(14.58%) | 2(16.67%) | 0.773 |
| Normal | 35(81.40%) | 36(75.00%) | 10(83.33%) | |
| regular | 3(6.98%) | 5(10.42%) | 0(0.00%) | |
| **Smoked Vegetables** | | | | |
| Never | 11(25.58%) | 14(29.17%) | 2(16.67%) | 0.733 |
| Normal | 29(67.44%) | 31(64.58%) | 8(66.67%) | |
| regular | 3(6.98%) | 3(6.25%) | 2(16.67%) | |
| **Salt Intake** | | | | |
| Less | 15(34.88%) | 15(31.25%) | 3(25.00%) | 0.800 |
| Heavy | 28(65.12%) | 33(68.75%) | 9(75.00%) | |
| **Oil Intake** | | | | |
| Less | 14(32.56%) | 15(31.25%) | 3(25.00%) | 0.882 |
| Heavy | 29(67.44%) | 33(68.75%) | 9(75.00%) | |
| **Water Intake** | | | | |
| >2 ltrs | 6(13.95%) | 7(14.58%) | 2(16.67%) | 0.985 |
| 1-2 ltrs | 9(20.93%) | 14(29.17%) | 3(25.00%) | |
| 500-1 ltr | 23(53.49%) | 22(45.83%) | 6(50.00%) | |

| | | | | |
|---|---|---|---|---|
| 1-2 glass | 5(11.63%) | 5(10.42%) | 1(8.33%) | |
| **Exercise** | | | | |
| never | 20(46.51%) | 27(56.25%) | 3 (25.00%) | 0.153 |
| 3-4 times | 10(23.26%) | 8(16.67%) | 6(50.00%) | |
| everyday | 13(30.23%) | 13(27.08%) | 3(25.00%) | |
| **Sleeping Hours** | | | | |
| 1-3 | 2(4.65%) | 4(8.33%) | 1 (8.33%) | 0.151 |
| 3-6 | 10(23.26%) | 18(37.50%) | 5(41.67%) | |
| 6-8 | 27(62.79%) | 21(43.75%) | 5(41.67%) | |
| 8-10 | 4(9.30%) | 5(10.42%) | 1(8.33%) | |
| **Night Duty** | | | | |
| yes | 9(20.93%) | 6(12.50%) | 3 (25.00%) | 0.438 |
| no | 34(79.07%) | 42(87.5%) | 9(75.00%) | |
| **Betel nut/pan** | | | | |
| Never | 8(18.60%) | 10(20.83%) | 5(41.67%) | 0.278 |
| Less | 14(32.56%) | 7(14.58%) | 2(16.67%) | |
| Normal | 10(23.26%) | 16(33.33%) | 3(25.00%) | |
| heavy | 11(25.58%) | 15(31.25%) | 2(16.67%) | |
| **Gutkha** | | | | |
| Never | 39(90.70%) | 44(91.67%) | 12(100.00%) | 0.978 |
| Less | 2(4.65%) | 2(4.17%) | 0(0.00%) | |
| Normal | 1(2.33%) | 1(2.08%) | 0(0.00%) | |
| heavy | 1(2.33%) | 1(2.08%) | 0(0.00%) | |
| **Sahdah** | | | | |
| Never | 13(30.23%) | 11(22.92%) | 6(50.00%) | 0.111 |
| Less | 11(25.58%) | 4(8.33%) | 1(8.33%) | |
| Normal | 6(13.95%) | 12(25.00%) | 2(16.67%) | |
| heavy | 13(30.23%) | 21(43.75%) | 3(25.00%) | |
| **Khaini** | | | | |
| Never | 33(76.74%) | 32(66.67%) | 10(83.33%) | 0.327 |
| Less | 4(9.30%) | 7(14.58%) | 0(0.00%) | |

| | | | | |
|---|---|---|---|---|
| Normal | 4(9.30%) | 2(4.17%) | 0(0.00%) | |
| heavy | 2(4.65%) | 7(14.58%) | 2(16.67%) | |
| **Tuibur** | | | | |
| Never | 32(74.42%) | 30(62.50%) | 6(50.00%) | |
| Less | 4(9.30%) | 10(20.83%) | 3(25.00%) | 0.593 |
| Normal | 5(11.63%) | 4(8.33%) | 2(16.67%) | |
| heavy | 2(4.65%) | 4(8.33%) | 1(8.33%) | |
| **Cigarette** | | | | |
| Never | 24(55.81%) | 34(70.83%) | 8(66.67%) | |
| Less | 6(13.95%) | 6(12.50%) | 1(8.33%) | 0.562 |
| Normal | 7(16.28%) | 2(4.17%) | 2(16.67%) | |
| heavy | 6(13.95%) | 6(12.50%) | 1(8.33%) | |
| **Alcohol** | | | | |
| Never | 41(95.35%) | 48(100.00%) | 12(100.00%) | 0.241 |
| Less | 2(4.65%) | 0(0.00%) | 0(0.00%) | |

**Table 14: Correlation between tumor grade lifestyle and dietary habits, tobacco and alcohol habits**

Analysis was performed among age group 40-55 years. However, no significant association was observed between tumor grades with the factors considered under lifestyle, dietary, tobacco and alcohol habits (Table 15).

| **Factors** | **Grade I (n= 21)** | **Grade II (n= 25)** | **Grade III (n= 5)** | **p-value** |
|---|---|---|---|---|
| **Pork** | | | | |
| Never | 2 (9.5%) | 3 (12.0%) | 0 (0.0%) | 0.858 |
| Once | 9 (42.9%) | 12 (48.0%) | 2 (40.0%) | |
| Thrice | 7 (33.3%) | 9 (36.0%) | 2 (40.0%) | |
| Everyday | 3 (14.3%) | 1 (4.0%) | 1 (20.0%) | |
| **Fish** | | | | |

| | | | | |
|---|---|---|---|---|
| Never | 0 (0.0%) | 3 (12.0%) | 0 (0.0%) | 0.511 |
| Once | 14 (66.7%) | 16 (64.0%) | 4 (80.0%) | |
| Thrice | 4 (19.0%) | 5 (20.0%) | 1 (20.0%) | |
| Everyday | 3 (14.3%) | 1 (4.0%) | 0 (0.0%) | |
| **Chicken** | | | | |
| Never | 1 (4.8%) | 1 (4.0%) | 0 (0.0%) | 0.881 |
| Once | 14 (66.7%) | 18 (72.0) | 3 (60.0%) | |
| Thrice | 3 (14.3%) | 5 (20.0%) | 1 (20.0%) | |
| Everyday | 3 (14.3%) | 1 (4.0%) | 1 (20.0%) | |
| **Beef** | | | | |
| Never | 2 (9.5%) | 4 (16.0%) | 0 (0.0%) | 0.642 |
| Once | 13 (61.9%) | 15 (60.0%) | 4 (80.0%) | |
| Thrice | 3 (14.3%) | 5 (20.0%) | 0 (0.0%) | |
| Everyday | 3 (14.3%) | 1 (4.0%) | 1 (20.0%) | |
| **Fruits** | | | | |
| Regular | 7 (33.3%) | 9 (36.0%) | 2 (40.0%) | 0.957 |
| Normal | 13 (61.9%) | 14 (56.0%) | 3 (60.0%) | |
| Never | 1 (4.8%) | 2 (8.0%) | 0 (0.0%) | |
| **Vegetables** | | | | |
| Regular | 17 (81.0%) | 18 (72.0%) | 5 (100.0%) | 0.476 |
| Normal | 4 (19.0%) | 5 (20.0%) | 0 (0.0%) | |
| Never | 0 (0.0%) | 2 (8.0%) | 0 (0.0%) | |
| **Sa-um** | | | | |
| Never | 6 (28.6%) | 5 (20.0%) | 2 (40.0%) | 0.884 |
| Normal | 9 (42.9%) | 13 (52.0%) | 2 (40.0%) | |
| Regular | 6 (28.6%) | 7 (28.0%) | 1 (20.0%) | |
| **Smoked meat** | | | | |
| Never | 3 (14.3%) | 5 (20.0%) | 1 (20.0%) | 0.693 |
| Normal | 13 (61.9%) | 13 (52.0%) | 4 (80.0%) | |
| Regular | 5 (23.8%) | 7 (28.0%) | 0 (0.0%) | |
| **Smoked vegetables** | | | | |

| | | | | |
|---|---|---|---|---|
| Never | 5 (23.8%) | 7 (28.0%) | 1 (20.0%) | 0.968 |
| Normal | 11 (52.4%) | 14 (56.0%) | 3 (60.0%) | |
| Regular | 5 (23.8%) | 4 (16.0%) | 1 (20.0%) | |
| **Salt intake** | | | | |
| Less | 10 (47.6%) | 5 (20.0%) | 2 (40.0%) | 0.133 |
| Heavy | 11 (52.4%) | 20 (80.0%) | 3 (60.0%) | |
| **Oil intake** | | | | |
| Less | 6 (28.6%) | 7 (28.0%) | 1 (20.0%) | 0.925 |
| Heavy | 15 (71.4%) | 18 (72.0%) | 4 (80.0%) | |
| **Water intake** | | | | |
| > 2 ltrs | 3 (14.3%) | 3 (12.0%) | 2 (40.0%) | 0.716 |
| 1-2 ltrs | 5 (23.8%) | 8 (32.0%) | 1 (20.0%) | |
| 500-1 ltr | 9 (42.9%) | 11 (44.0%) | 2 (40.0%) | |
| 1-2 glass | 4 (19.0%) | 3 (12.0%) | 0 (0.0%) | |
| **Exercise** | | | | |
| Everyday | 8 (38.1%) | 8 (32.0%) | 1 (20.0%) | 0.752 |
| 3-4 times | 4 (19.0%) | 4 (16.0%) | 2 (40.0%) | |
| Never | 9 (42.9%) | 13 (52.0%) | 2 (40.0%) | |
| **Sleeping hours** | | | | |
| 8-10 hrs | 2 (9.5%) | 2 (8.0%) | 1 (20.0%) | 0.823 |
| 6-8 hrs | 14 (66.7%) | 13 (52.0%) | 2 (40.0%) | |
| 3-6 hrs | 4 (19.0%) | 9 (36.0%) | 2 (40.0%) | |
| 1-3 hrs | 1 (4.8%) | 1 (4.0%) | 0 (0.0%) | |
| **Night duty** | | | | |
| Yes | 16 (76.2%) | 22 (88.0%) | 5 (100.0%) | 0.327 |
| No | 5 (23.8%) | 3 (12.0%) | 0 (0.0%) | |
| **Betel nut/pan** | | | | |
| Never | 4 (19.0%) | 6 (24.0%) | 2 (40.0%) | |
| Less | 6 (28.6%) | 3 (12.0%) | 1 (20.0%) | 0.814 |
| Normal | 5 (23.8%) | 8 (32.0%) | 1 (20.0%) | |
| Heavy | 6 (28.6%) | 8 (32.0%) | 1 (20.0%) | |

| | | | | |
|---|---|---|---|---|
| **Gutkha** | | | | |
| Never | 20 (95.2%) | 22 (88.0%) | 5 (100.0%) | 0.796 |
| Less | 1 (4.8%) | 2 (8.0%) | 0 (0.0%) | |
| Normal | 0 (0.0%) | 1 (4.0%) | 0 (0.0%) | |
| **Sahdah** | | | | |
| Never | 5 (23.8%) | 6 (24.0%) | 2 (40.0%) | 0.174 |
| Less | 6 (28.6%) | 1 (4.0%) | 0 (0.0%) | |
| Normal | 4 (19.0%) | 6 (24.0%) | 0 (0.0%) | |
| Heavy | 6 (28.6%) | 12 (48.0%) | 3 (60.0%) | |
| **Khaini** | | | | |
| Never | 16 (76.2%) | 14 (56.0%) | 4 (80.0%) | 0.605 |
| Less | 2 (9.5%) | 5 (20.05) | 0 (0.0%) | |
| Normal | 2 (9.5%) | 2 (8.0%) | 0 (0.0%) | |
| Heavy | 1 (4.8%) | 4 (16.0%) | 1 (20.0%) | |
| **Tuibur** | | | | |
| Never | 14 (66.7%) | 17 (68.0%) | 3 (60.0%) | 0.437 |
| Less | 2 (9.5%) | 5 (20.0%) | 2 (40.0%) | |
| Normal | 3 (14.3%) | 3 (12.0%) | 0 (0.0%) | |
| Heavy | 2 (9.5%) | 0 (0.0%) | 0 (0.0%) | |
| **Cigarette** | | | | |
| Never | 14 (66.7%) | 18 (72.0%) | 4 (80.0%) | 0.798 |
| Less | 2 (9.5%) | 3 (12.0%) | 0 (0.0%) | |
| Normal | 3 (14.3%) | 1 (4.0%) | 0 (0.0%) | |
| Heavy | 2 (9.5%) | 3 (12.0%) | 1 (20.0%) | |
| **Alcohol** | | | | |
| Never | 20 (95.2%) | 25 (100.0%) | 5 (100.0%) | 0.483 |
| Less | 1 (4.8%) | 0 (0.0%) | 0 (0.0%) | |

**Table 15: Correlation between tumor grade with lifestyle, dietary, tobacco and alcohol habits among patients diagnosed at age between 40-55 years.**

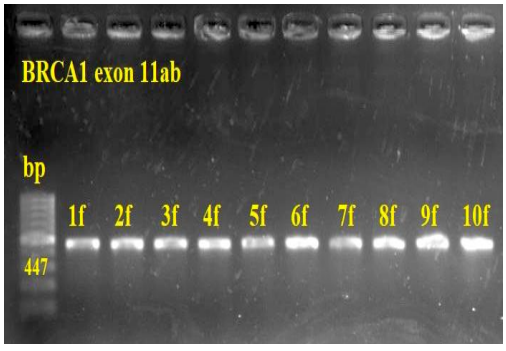
***3. Study of mutations in candidate genes and their association with BC in Mizo Population***

***3.1. Study of genetic alteration using Direct Sanger Sequencing*:**

Pictures of agarose gel showing isolated genomic DNA is given in figure 5.


**Figure 5: Representative gel picture of isolated genomic DNA from blood samples (**F- Familial Breast Cancer).

Amplification of selected genes such as *BRCA1, TP53, PTEN, CDH1, CHEK2, XRCC2* was carried out using PCR for the specified region of interest [Figure 6 (i-vi)].



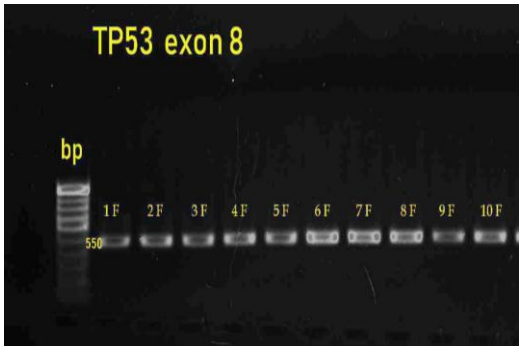
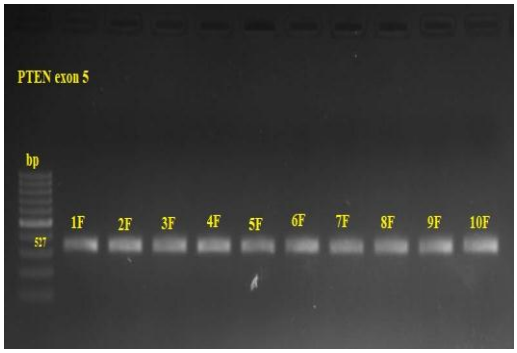




**Figure 6(i): *BRCA1* amplified products.** **Figure 6(ii): *TP53* amplified products.**



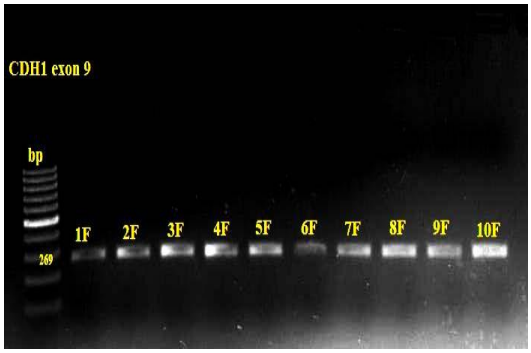
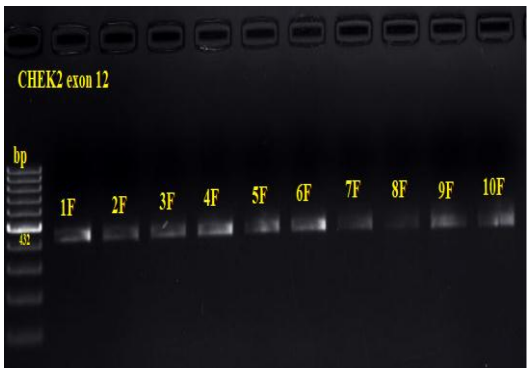




**Figure 6(iii): *PTEN* amplified products.** **Figure 6(iv): *CDH1* amplified products**



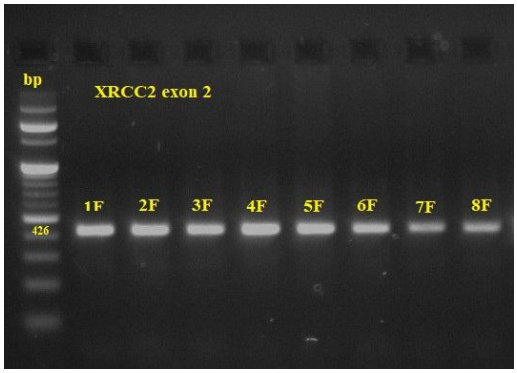




**Figure 6(v): *CHEK2* amplified products.** **Figure 6(vi): *XRCC2* amplified products**

**Figure 6: Amplification showing the estimated product size of the selected region of interest. i). BRCA1 ii). TP53 iii). PTEN iv). CDHI v). CHEK2 vi). XRCC2**

From this analysis, eight polymorphisms in three exons (11, 13 and 15) of *BRCA1* gene was identified. Four are non- synonymous and the other four are synonymous. Among the synonymous polymorphisms, one variant, g.95900A>T: c.4772A>T: p.P1544P in exon 15 was found to be novel. Several genetics clinics and research groups had reported the other seven remaining polymorphisms in BIC, ARUP BRCA, dbSNP and ClinVar databases. Details of the polymorphisms and chromatographic files are given in Table 16 and Figure 7 (i to vii), respectively. There was no significant amino acid change to be noted since all the polymorphisms represent silent substitutions. These polymorphisms were observed in more than 90% of the samples (in both cases and controls). In case of *TP53, PTEN, CDH1, CHEK2* and *XRCC2* gene, no genetic alterations in the exons of interest was observed.

| Exon | Genomic Position (Ensembl NM_007294) | HGVS | A.A change (HGVS) | rs ID | Align GVGD Score | Zygosity | ExAC |
|---|---|---|---|---|---|---|---|
| 11 | g.77054T>C: c.2451T>C | g.41245237A>G | p.L771L | 16940 | C0 | Homo/ Hetero | 1. African 2.South Asian 3.European (Finnish & NonFinnish) 4.East asian 5.Latino |
| | g.76825C>T: c.2222C>T | g.41245466G>A | p.S694S | 1799949 | C0 | Homo | |
| | g.77355C>T: c.2752C>T | g.41244936G>A | p.P871L | 799917 | C0 | Homo/ Hetero | |
| | g.77856A>G: c.3253A>G | g.41244435T>C | p.E1038G | 16941 | C0 | Homo/ Hetero | |
| | g.78291A>G: c.3688A>G | g.41244000T>C | p.K1183R | 16942 | C0 | Homo/ Hetero | |
| 13 | g.87821T>C: c.4448T>C | g.41234470A>G | p.S1436S | 1060915 | C0 | Homo/ Hetero | |
| 15 | g.95900A>T: c.4772A>T | novel | p.P1544P | novel | C0 | Hetero | |
| 15 | g.95904G>T: c.4776G>T | g.41226387C>A | p.D1546Y | 28897691 | C0 | Hetero | |

**Table 16: Polymorphisms observed in BRCA1 gene**

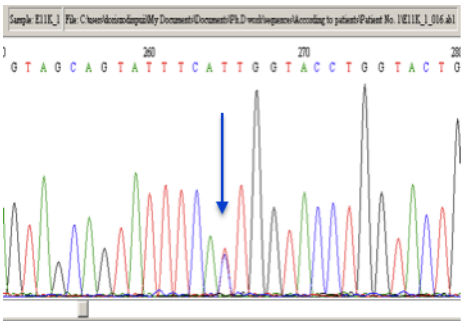


**Figure 7: Exon-11, g.77054T>C**

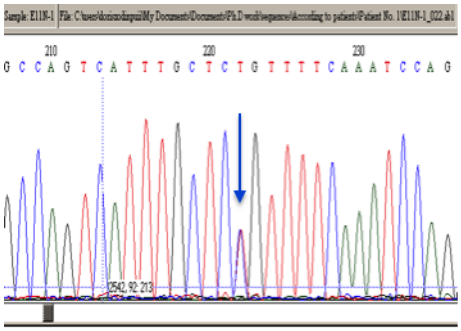


**Figure 8: Exon-11, g.77355C>T**

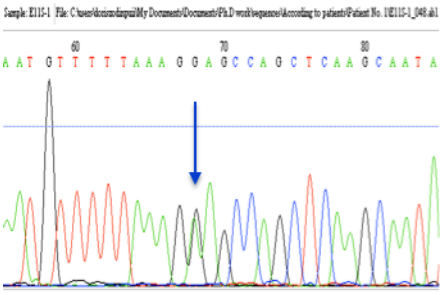


**Figure 9: Exon-11, g.77856A>G**

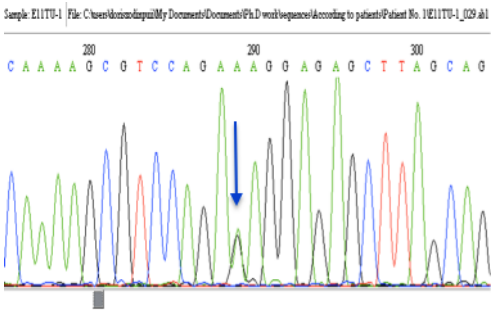


**Figure 10: Exon-11, g.78291A>G**

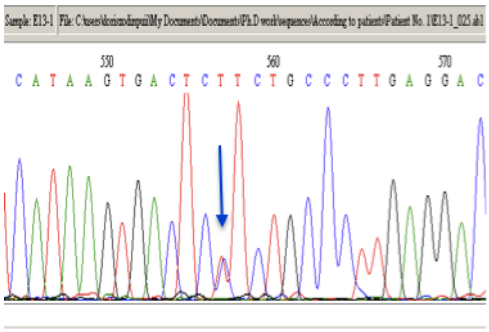


**Figure 11: Exon-13, g.87821T>C**

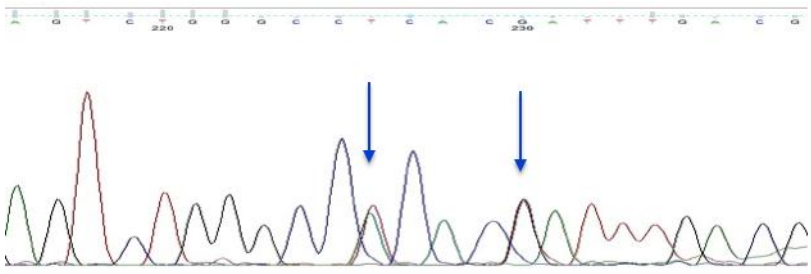


**Figure 12: Exon-15, g.95900A>T and g.95904G>T**

## 2.2. Study of BC susceptibility gene BRCA1 and BRCA2 using Targeted re-sequencing

In this *BRCA1* and *BRCA2* whole gene analysis consisting of 15 samples (10 BC cases and 5 healthy controls), two variants (c.T5089C:p.C1697R and c.C376T:p.Q126X) in *BRCA1* gene from sample 1F and 5F, respectively were found to be pathogenic. These two observed variants have been reported in databases for other populations and c.T5089C:p.C1697R is predicted to have a high score of pathogenicity according to ALIGN- GVGD grading. Among the studied samples, we did not observe any pathogenic variants in *BRCA2* gene. Observed mutation summary and chromatogram is given in Figures 13, 14 and Table 17.

Structural prediction was performed for the altered variants, however we could obtain only for missense mutation (p.C1697R) that falls on exon 18 (Figure 15) since the prediction tool could only solve missense variants. This nonsense pathogenic variant denoted as BRCA1 c.376C>T at the cDNA level and p.Gln126Ter (Q126X) at the protein level changes a Glutamine to a premature stop codon (CAA>TAA), and is predicted to cause loss of normal protein function through either protein truncation or nonsense-mediated mRNA decay.

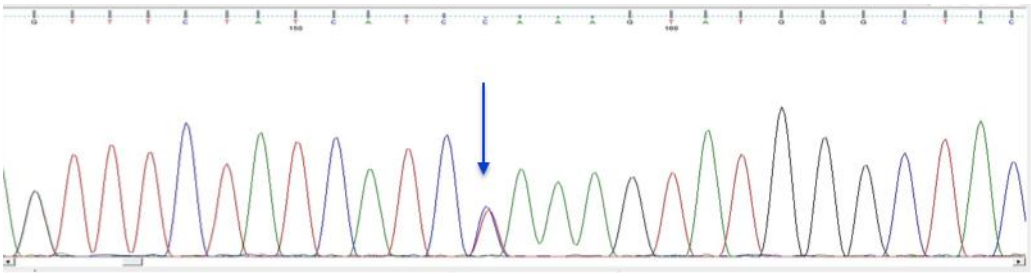


**Figure 13: BRCA1 exon 18 g.41215954 c.T5089C p.C1697R**

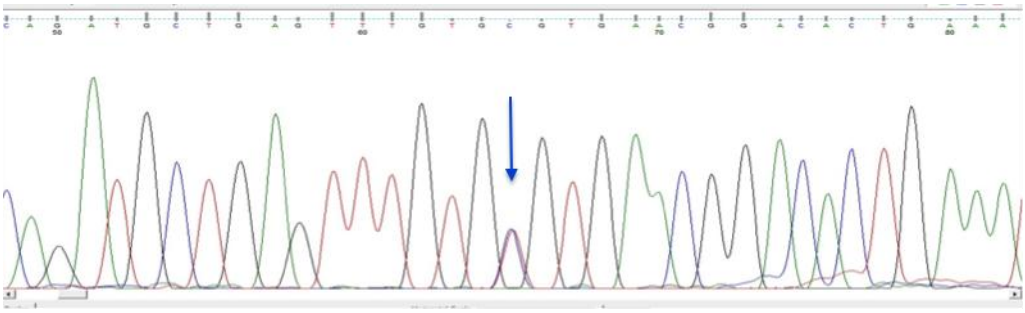


**Figure 14: BRCA1 exon 7  g.41256204 c.C376T p.Q126X**

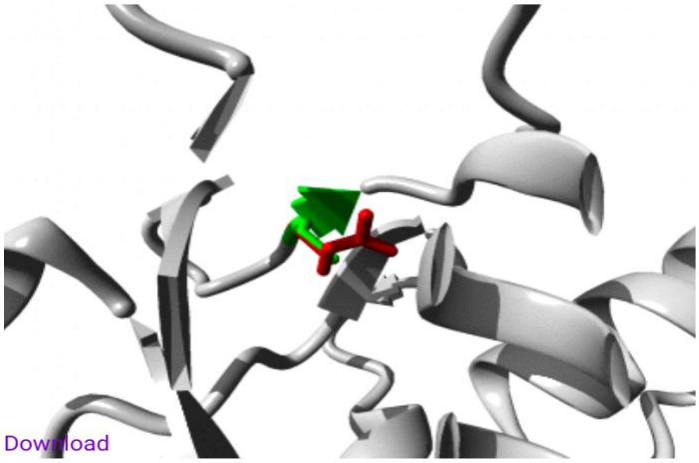



**Figure 15:  Structural prediction for p.C1697R**

The proteins are represented in grey color, wildtype side chains in green color and mutant residues in red color.

| **Patient ID** | **Exon** | **Position** | **cDNA & AA. Chg.** | **rs ID** | **Mutation type** | **Classification** | **Posterior probability** | **Align GVGD** |
|---|---|---|---|---|---|---|---|---|
| 1F | 18 | g.41215954 | c.T5089C p.C1697R | 80356993 | Missense | Likely pathogenic | 0.81 | Grade C65 |
| 5F | 7 | g.41256204 | c.C376T p.Q126X | nonsense | nonsense | 5-Definitely Pathogenic | >0.99 | - |

**Table 17: Pathogenic variants observed in BRCA1 gene, exon 18 and 7.**

**2.3. Study on clinically significant mutations using Clinical Exome sequencing**

From this study comprising of 12 samples (7 familial BC cases, 2 non-familial BC cases and 3 healthy controls), 16 variants that meets all the filtered parameters were used for the analysis. According to the ACMG (American College of Medical Genetics) variants classification guidelines, the 16 variants observed consisted of 5- benign variants, 3- either benign or VUS (Variants of Uncertain Significance), 1- likely benign and 7 variants falls into VUS. Also, from the observed variants, 3 variants in *BLM* (p.Thr162Ile), *CDKN2A* (intronic) and *FANCC* (p.Asp197Gly) were found to be novel (Table 18). Protein structural prediction was performed for novel missense variants, however, we could not obtain a structural information for *BLM* c.485C>T: p.Thr162Ile and *FANCC* c.590A>G: p.Asp197Gly since the 3D-structure was not available. Only structures of the amino acid residue change (wild type and mutant residue) could be obtain for both the substitution (Figures 16 and 17, respectively).

| GENE | GENOMIC POSITION | VARIANT CLASS | cDNA: AA_CHG | MAF (ExAC) | SAMPLE | CLASSIFICATION (ACMG) |
|---|---|---|---|---|---|---|
| BARD1 | 215645502 | inframe-del | c.1075_1095del: p.Leu359_Pro365del | 0.029 | 27f, 41f | Benign |
| BLM | 91292983 | missense | c.485C>T: p.Thr162Ile | Not reported | 41F | Vus |
| BLM | 91295110 | missense | c.893C>T: p.Thr298Met | 0.008 | 26f, 27f | Benign |
| CDH1 | 68867391 | missense | c.2638G>A: p.Glu880Lys | 0.000 | 26F | Benign/ Vus |
| CDKN2A | 21968771 | intronic-ss-acr | c.458-1G>A: NA | Not reported | 8F | Vus |
| FANCC | 97912301 | missense | c.590A>G: p.Asp197Gly | Not reported | 51NF | Vus |
| FANCM | 45644706 | missense | c.2749A>G: p.Ile917Val | 0.004 | 8f, 51nf | Benign |
| GEN1 | 17947927 | missense | c.607A>G: p.Ile203Val | 0.016 | 31F | Vus |
| GEN1 | 17963098 | missense | c.2619T>G: p.Ser873Arg | 0.006 | 31F | Vus |

| | | | | | | |
|---|---|---|---|---|---|---|
| GEN1 | 17942856 | intronic-ss-dnr | c.348+7A>G: NA | 0.016 | 31F | Vus |
| MSH6 | 48033981 | frameshift-ins | c.4068_4071dup p.Lys1358AspfsTer2 | 0.002 | 27F | likely benign |
| MUTYH | 45797401 | missense | c.1109C>T: p.Ala370Val | 0.000385021 | 27f, 51nf | Vus |
| PALB2 | 23646375 | missense | c.1492G>T: p.Asp498Tyr | 0.000510708 | 27F | Benign/Vus |
| PMS2 | 6026864 | missense | c.1532C>T: p.Thr511Met | 0.0093915 | 45F | Benign |
| STK11 | 1223125 | missense | c.1062C>G: p.Phe354Leu | 0.00575723 | 51nf, 147nf | Benign/Vus |
| VHL | 10183876 | intronic-ss-dnr-prx | c.340+5G>C: NA | 0.0145216 | 31F | Benign |

**Table 18: Variants observed from clinical exome sequencing using ACMG guidelines**.


**Figure 16: Structural prediction for BLM c.485C>T: p.Thr162Ile.**

Left- wildtype and right- mutant. The backbone, which is similar for both amino acid, is shown in red color. The side chain, unique for each amino acid, is shown in black color.


**Figure 17: Structural prediction for FANCC c.590A>G: p.Asp197Gly.**

Left- wildtype and right- mutant. The backbone, which is similar for both amino acid, is shown in red color. The side chain, unique for each amino acid, is shown in black color.

The variants found through the clinical exome sequencing were analyzed using the software prediction tools such as SIFT, Polyphen2, Provean, LRT, FATHMM, METASVM, METALR and CADD fetched us 12 deleterious variants. 5 among the observed variants were also found in the above approached (*CDH1*:p.Gly62Ser, *CDKN2A*, *MSH6*: p.Lys1358AspfsTer2, *MUTYH:* p.Ala370Val, *PALB2:* p.Asp498Tyr). Majority of the variants are missense while only two variants are intronic-ss-acr and frameshift-ins (Table 19).

| GENE | GENOMIC POSITION | CDNA: AA_CHG | VARIANT CLASS | SOFTWARE PREDICTION | NOVELTY STATUS |
|---|---|---|---|---|---|
| ATM | 108186820 | c.6178C>T p.Arg2060Cys | missense | Deleterious | Reported |
| BRCA2 | 32906480 | c.865A>C p.Asn289His | missense | Deleterious | Reported |
| CDH1 | 68867391 | c.2638G>A p.Glu880Lys | missense | Deleterious | Reported |
| CDH1 | 68835593 | c.184G>A p.Gly62Ser | missense | Deleterious | Reported |
| CDKN2A | 21968771 | c.458-1G>A NA | intronic-ss-acr | Deleterious | Not Reported |
| MRE11A | 94219145 | c.259C>T p.Arg87Trp | missense | Deleterious | Reported |
| MSH6 | 48033981 | c.4068_4071dup p.Lys1358AspfsTer2 | frameshift-ins | Deleterious | Reported |
| MUTYH | 45797401 | c.1109C>T p.Ala370Val | missense | Deleterious | Reported |
| MUTYH | 45797914 | c.848G>A p.Gly283Glu | missense | Deleterious | Reported |
| PALB2 | 23646375 | c.1492G>T p.Asp498Tyr | missense | Deleterious | Reported |
| STK11 | 1223151 | c.1088C>T p.Thr363Ile | missense | Deleterious | Reported |
| TP53 | 7578388 | c.542G>A p.Arg181His | missense | Deleterious | Reported |

**Table 19: Variants observed based on software prediction tools**

Structural prediction was obtained for the pathogenic variants *MUTYH* (c.848G>A: p.Gly283Glu) and *TP53* (c.542G>A: p.Arg181His) in which the proteins are shown in grey color and the side chains of both the wildtype and mutant residues are shown in green and red color, respectively (Figures 18 and 19).

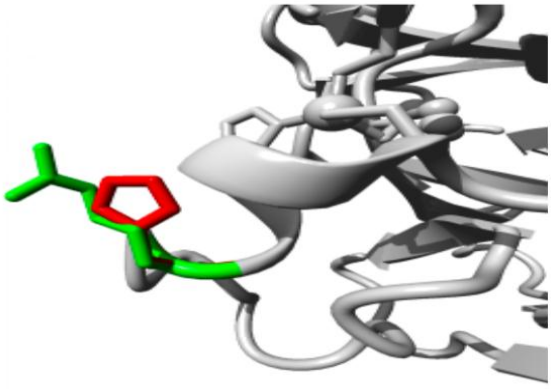


**Figure 18: Structural prediction of TP53 (c.542G>A: p.Arg181His)**

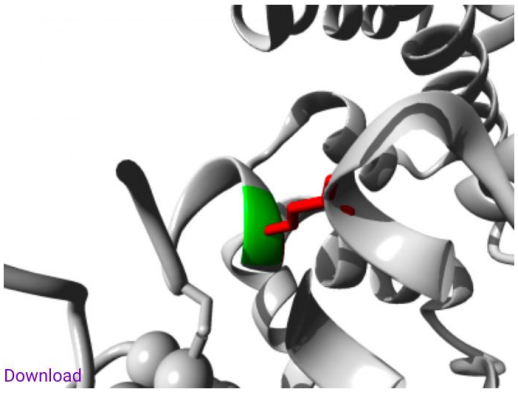



**Figure 19: Structural prediction of MUTYH (c.848G>A: p.Gly283Glu)**

### 2.4. Study on mitochondrial gene alteration in BC using whole mitochondrial sequencing

In mtDNA genome, sequence alteration was found to be the highest in d-loop region and least in ND6 region (Figure 20).

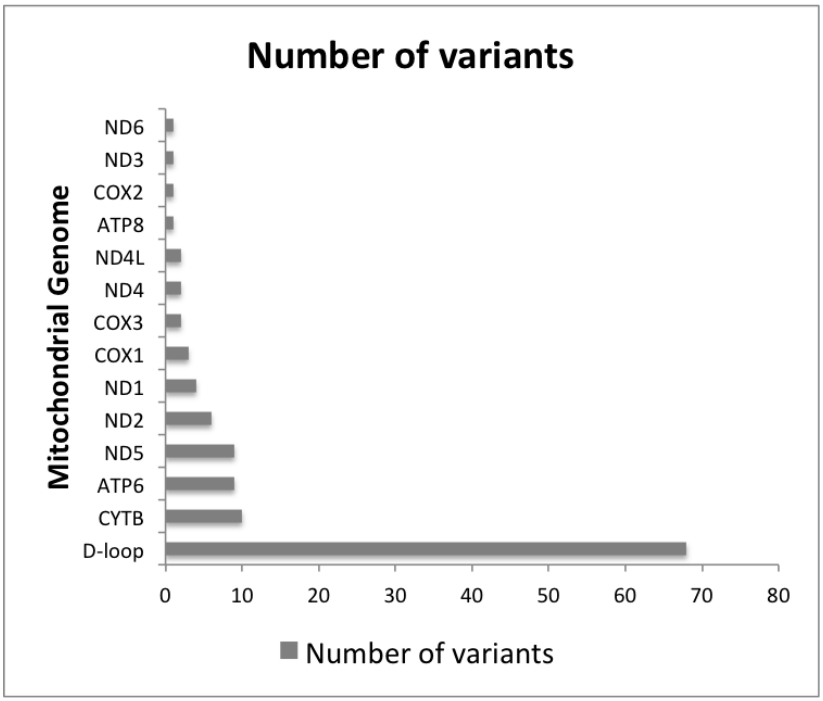



**Figure 20: Frequency distribution of variants in mitochondrial genome.**

y-axis- Coding genes and non coding d-loop region of mtDNA

x-axis- Number of alteration in each region

A total of 119 non-synonymous variants were found in the samples studied (cases and controls). Among them, 51 variants were observed in the protein coding region of mtDNA, in which 27 were specific to BC cases and 24 were found in common among the cases and controls. One frameshift variants (12014C>CA; c.1255_1255delinsCA) located in *ND4* gene was found to be novel. Four variants (*ATP6* 8860A>G, *ND3* 10398A>G, *ND5* 13708G>A and *ND5* 14110T>C) found in

common for both cases and controls have been reported in database for BC. According to SIFT and Polyphen- 2, 22 variants were found to be affecting the normal protein function. Based on SOSUI software, 98% of non- synonymous variants falls in the hydrophobic region. The potential impact of a greater number of non-synonymous substitutions in the protein coding region represented decrease in protein stability, whereas, only two variant in ND2 and ND6 with ΔΔG of 0.144 and 0.025, respectively, were observed with increase protein stability according to Pmut (Table 20).

| Gene | Position | AA. change | Freq. (%) in cases/control | SIFT | Polyphen 2 Score | SOSUI | PMUT SCORE delta delta G | |
|---|---|---|---|---|---|---|---|---|
| ATP6 | 8686T>C | c.T160C:p.S54P | 3.12 | T | PD | 0.94 | -0.85 | BC specific |
| | 8869A>C | c.A343C:p.M115L | 3.12 | T | B | 0.96 | -0.45 | |
| | 9053G>A | c.G527A:p.S176N | 3.12 | T | B | 0.93 | -0.84 | |
| COX1 | 6120A>G | c.A217G:p.I73V | 3.12 | T | PD | 0.68 | -0.72 | |
| | 6340C>T | c.C437T:p.T146I | 3.12 | APF | B | 0.69 | -0.48 | |
| | 6480G>A | c.G577A:p.V193I | 3.12 | T | B | 0.68 | -0.76 | |
| COX2 | 7853G>A | c.G268A:p.V90I | 3.12 | T | B | 0.46 | -0.84 | |
| COX3 | 9469C>T | c.C263T:p.T88I | 3.12 | T | B | 0.39 | -0.46 | |
| | 9966G>A | c.G760A:p.V254I | 3.12 | T | B | 0.37 | -0.39 | |
| CYTB | 14793A>G | c.A47G:p.H16R | 3.12 | APF | B | 0.62 | -1.10 | |
| | 15497G>A | c.G751A:p.G251S | 3.12 | APF | B | 0.62 | -1.46 | |
| | 15644A>G | c.A898G:p.I300V | 3.12 | T | PD | 0.63 | -0.80 | |
| | 15884G>A | c.G1138A:p.A380T | 6.25 | T | PD | 0.62 | -1.01 | |
| ND1 | 3398 T>C | c.T92C:p.M31T | 3.12 | APF | B | 0.67 | -1.58 | |
| | 3505A>G | c.A199G:p.T67A | 3.12 | APF | B | 0.68 | -0.91 | |
| | 3578 T>C | c.T272C:p.M91T | 3.12 | APF | PD | 0.67 | -0.85 | |
| ND2 | 4596G>A | c.G127A:p.V43I | 3.12 | T | B | 0.63 | 0.14 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 4833A>G | c.A364G:p.T122A | 6.25 | APF | PD | 0.64 | -0.54 | |
| ND4 | 12026A>G | c.A1267G:p.I423V | 3.12 | T | B | 0.72 | -0.67 | |
| | 12014C>CA | c.1255_1255delinsCA | 12.5 | - | - | - | Not reported | |
| ND5 | 13105A>G | c.A769G:p.I257V | 3.12 | T | B | 0.58 | -0.66 | |
| | 13153A>G | c.A817G:p.I273V | 3.12 | T | B | 0.58 | -0.94 | |
| | 13759G>A | c.G1423A:p.A475T | 3.12 | T | B | 0.58 | -0.66 | |
| | 14002A>G | c.A1666G:p.T556A | 3.12 | T | B | 0.59 | -1.53 | |
| | 14110 T>C | c.T1774C:p.F592L | 3.12 | T | B | 0.59 | -1.30 | |
| | 12417 C>-A | c.81_81delins-A | 9.37 | - | - | - | frameshift | |
| ND6 | 14553C>T | c.G121A:p.V41I | 3.12 | APF | B | 1.07 | 0.025 | |
| ATP6 | 8584 G>A | c.G58A:p.A20T | 6.25/ 8.33 | T | B | 0.94 | -1.83 | |
| | 8602 T>C | c.T76C:p.F26L | 3.12/ 16.66 | T | B | 0.95 | -0.63 | |
| | 8701 A>G | c.A175G:p.T59A | 3.12/ 33.33 | T | B | 0.96 | -0.78 | |
| | 8794 C>T | c.C268T:p.H90Y | 3.12/ 8.33 | T | B | 0.96 | -0.03 | |
| | 8860 A>G | c.A334G:p.T112A | 3.12/ 100 | T | B | 0.96 | -0.39 | |
| | 9094 C>T | c.C568T:p.L190F | 3.12/ 8.33 | T | PrD | 0.94 | -0.71 | |
| ATP8 | 8414 C>T | c.C49T:p.L17F | 9.37/ 8.33 | T | PrD | -0.37 | -1.46 | |
| CYTB | 14766 C>T | c.C20T:p.T7I | 100/ 100 | APF | B | 0.64 | -0.14 | |
| | 15047 G>A | c.G301A:p.G101S | 6.25/ 16.66 | APF | B | 0.62 | -0.76 | |

| | | | | | | | | |
|---|---|---|---|---|---|---|---|---|
| | 15326 A>G | c.A580G:p.T194A | 100/ 100 | APF | B | 0.63 | -0.42 | Common variants |
| | 15402 C>T | c.C656T:p.T219I | 25/ 16.66 | APF | PD | 0.64 | -0.35 | |
| | 15449 T>C | c.T703C:p.F235L | 3.12/ 8.33 | T | B | 0.63 | -1.44 | |
| | 15825 C>T | c.C1079T:p.T360M | 6.25/ 16.66 | T | PD | 0.63 | -0.42 | |
| ND1 | 3316 G>A | c.G10A:p.A4T | 3.12/ 8.33 | T | B | 0.67 | -1.12 | |
| ND2 | 4824 A>G | c.A355G:p.T119A | 3.12/ 8.33 | T | PD | 0.64 | -0.39 | |
| | 5178 C>A | c.C709A:p.L237M | 9.37/ 8.33 | T | PrD | 0.63 | -0.90 | |
| | 5186 A>T | c.A717T:p.W239C | 3.12/ 8.33 | APF | PrD | 0.64 | -0.84 | |
| | 5460 G>A | c.G991A:p.A331T | 3.12/ 16.66 | T | B | 0.62 | -1.24 | |
| ND3 | 10398 A>G | c.A340G:p.T114A | 3.12/ 41.66 | T | B | 1.01 | -0.92 | |
| ND4L | 10609 T>C | c.T140C:p.M47T, | 3.12/ 25 | T | B | 1.26 | -1.27 | |
| | 10653 G>A | c.G184A:p.A62T | 6.25/ 8.33 | T | B | 1.26 | -1.18 | |
| ND5 | 12406 G>A | c.G70A:p.V24I | 31.25/ 25 | T | B | 0.58 | -0.19 | |
| | 13708 G>A | c.G1372A:p.A458T | 9.37/ 16.66 | T | B | 0.58 | -1.17 | |
| | 13928 G>C | c.G1592C:p.S531T | 31.25/ 33.33 | T | PD | 0.58 | -0.69 | |

**Table 20: Non- synonymous variants found in the mtDNA coding region.**

APF - Affect Protein Function, T- Tolerated, PD – Possibly Damaging, PRD – Probably Damaging, B- Benign

In total, 68 variants in the non-coding displacement loop (d- loop) region of mitochondrial genome was observed, comprising of 26 BC specific variants and 42 common variants. The variant specific to BC cases, 16126T>C and other eight common variants such as 73A>G, 152T>C, 263A>G, 16189T>C, 16223C>T, 16290C>T, 16304T>C, 16362T>C has been reported for BC. Other populations have reported the remaining variants for other diseases, but not for BC (Table 21).

| | **Position** | **Freq.(%)** | | **Position** | **Freq.(%)** |
|---|---|---|---|---|---|
| BC specific | 207 G>A | 6.25 | Common Variants | 189 A>G | 3.12/ 8.33 |
| | 279 T>C | 3.12 | | 195 T>C | 9.37/ 16.66 |
| | 302 A>ACC | 3.12 | | 200 A>G | 6.25/ 8.33 |
| | 16111 C>T | 3.12 | | 214 A>G | 3.12/ 8.33 |
| | 16126 T>C | 3.12 | | 234 A>G | 6.25/ 8.33 |
| | 16136 T>C | 3.12 | | 235 A>G | 3.12/ 8.33 |
| | 16140 T>C | 3.12 | | 247 G>-A | 15.62/ 33.33 |
| | 16169 C>T | 3.12 | | 263 A>G | 93.75/ 37.5 |
| | 16179 C>-A | 6.25 | | 302 A>AC | 28.12/ 33.33 |
| | 16182 A>AC | 6.25 | | 310 T>TC | 15.62/ 100 |
| | 16183 A>ACC | 15.62 | | 16051 A>G | 6.25/ 8.33 |
| | 16183 A>C | 6.25 | | 16092 T>C | 12.5/ 16.66 |
| | 16214 C>T | 3.12 | | 16129 G>A | 6.25/ 8.33 |
| | 16220 A>G | 3.12 | | 16183 A>AC | 25/ 25 |
| | 16227 A>G | 6.25 | | 16185 C>T | 3.12/ 8.33 |
| | 16248 C>T | 3.12 | | 16188 C>-T | 6.25/ 33.33 |

| | | | | | |
|---|---|---|---|---|---|
| | 16249 T>C | 3.12 | | 16189 T>C | 34.37/ 25 |
| | 16258 A>C | 3.12 | | 16209 T>C | 9.37/ 8.33 |
| | 16266 C>T | 6.25 | | 16213 G>A | 9.37/ 8.33 |
| | 16272 A>G | 3.12 | | 16218 C>T | 6.25/ 8.33 |
| | 16278 C>T | 6.25 | | 16223 C>T | 50/ 18.75 |
| | 16288 T>C | 3.12 | | 16234 C>T | 3.12/ 8.33 |
| | 16324 T>C | 6.25 | | 16239 C>T | 3.12/ 8.33 |
| | 16335 A>G | 3.12 | | 16246 A>C | 9.37/ 8.33 |
| | 16355 C>T | 6.25 | | 16260 C>T | 6.25/ 8.33 |
| | 16356 T>C | 3.12 | | 16289 A>G | 6.25/ 16.66 |
| Common Variants | 73 A>G | 100/ 100 | | 16290 C>T | 3.12/ 8.33 |
| | 94 G>A | 3.12/ 8.33 | | 16298 T>C | 6.25/ 8.33 |
| | 146 T>C | 37.5/ 25 | | 16304 T>C | 40.62/ 15.62 |
| | 151 C>T | 6.25/ 8.33 | | 16311 T>C | 43.75/ 21.9 |
| | 152 T>C | 34.37/ 21.9 | | 16319 G>A | 3.12/ 8.33 |
| | 183 A>G | 3.12/ 16.66 | | 16352 T>C | 3.12/ 8.33 |
| | 184 G>A | 3.12/ 16.66 | | 16353 C>T | 3.12/ 8.33 |
| | 185 G>A | 6.25/ 25 | | 16362 T>C | 37.5/ 33.33 |

**Table 21: Non- synonymous variants found in mtDNA D-loop region**

The ratio of non-synonymous to synonymous variants for mitochondrial respiratory complexes was analyzed and it was observed that the complexes vary in ratio. Complex V with 73% has the highest ratio in comparison to complex I, III and IV that were all lower than 40% (Figure 21). The dN/dS ratio of the variants in complex I, III, IV and V were 3.24, 0.894, 1.71 and 0.164, respectively. Since complex I and III were found to exceed 1, it indicates that there is a positive selection for non- synonymous variants in the two complexes in BC samples.

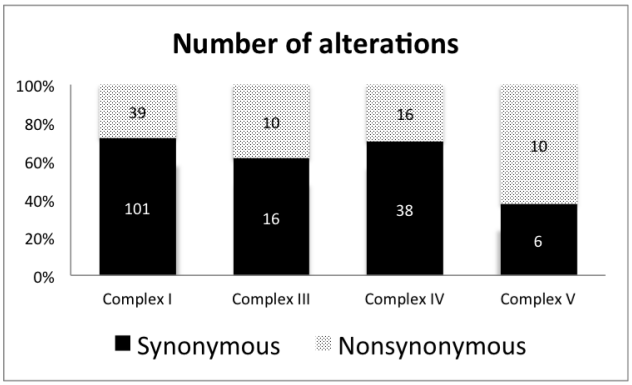



**Figure 21: Distribution of Synonymous and Non- synonymous variants in breast cancer cases**. (Complex I: ND1 to ND6; Complex III: CYTB; Complex IV: COX1, COX2, COX3; Complex V: ATP6, ATP8).

From the correlation analysis of 66 common variants present in Mitochondrial genome, the frequencies of four variants were found to be significantly different between the cases and controls. Three of them were located in the coding region (ATP6 8701A>G, ND3 10398A>G and ND4L 10609T>C) and one in the d-loop region (16188 C>-T). These variants obtain a high odds ratio (> 1), which signifies that they are high risk factors for BC occurrence (Table 22).

| Gene | position | Controls % | Cases % | p-value | OR | 95% CI |
|---|---|---|---|---|---|---|
| ATP6 | 8584 G>A | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 8602 T>C | 4.5 | 2.3 | 0.17 | 6.200 | 0.507-75.838 |
| | 8701 A>G* | 9.1 | 2.3 | 0.018 | 15.500 | 1.516-158.524 |
| | 8794 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 8860 A>G | 27.3 | 2.3 | 0.00 | - | - |
| | 9094 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| ATP8 | 8414 C>T | 2.3 | 6.8 | 1.00 | 0.879 | 0.082-9.374 |
| CYTB | 14766 C>T | 27.3 | 72.7 | - | | |
| | 15047 G>A | 4.5 | 4.5 | 0.29 | 3.000 | 0.372-24.171 |
| | 15326 A>G | 27.3 | 72.7 | - | | |
| | 15402 C>T | 4.5 | 18.2 | 0.70 | 0.600 | 0.108-3.338 |
| | 15449 T>C | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 15825 C>T | 4.5 | 4.5 | 0.29 | 3.000 | 0.372-24.171 |
| ND1 | 3316 G>A | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| ND2 | 4824 A>G | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 5178 C>A | 2.3 | 6.8 | 1.00 | 0.879 | 0.082-9.374 |
| | 5186 A>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 5460 G>A | 4.5 | 2.3 | 0.17 | 6.200 | 0.507-75.838 |
| ND3 | 10398 A>G* | 11.4 | 2.3 | 0.004 | 22.143 | 2.224-220.494 |
| ND4L | 10609 T>C* | 6.8 | 2.3 | 0.05 | 10.333 | 0.955-111.845 |
| ND4L | 10653 G>A | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| ND5 | 12406 G>A | 6.8 | 22.7 | 1.00 | 0.733 | 0.163-3.304 |
| | 13708 G>A | 4.5 | 6.8 | 0.60 | 1.933 | 0.281-13.295 |
| | 13928 G>C | 9.1 | 22.7 | 1.00 | 1.100 | 0.267-4.523 |

| | | | | | | |
|---|---|---|---|---|---|---|
| D-loop | 16051 A>G | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 16092 T>C | 4.5 | 9.1 | 0.65 | 1.400 | 0.221-8.856 |
| | 16129 G>A | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 16183 A>AC | 6.8 | 18.2 | 1.00 | 1.000 | 0.216-4.628 |
| | 16185 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16188 C>-T* | 9.1 | 4.5 | 0.03 | 7.500 | 1.158-48.564 |
| | 16189 T>C | 6.8 | 25.0 | 0.72 | 0.636 | 0.142-2.842 |
| | 16209 T>C | 2.3 | 6.8 | 1.00 | 0.879 | 0.082-9.374 |
| | 16213 G>A | 2.3 | 6.8 | 1.00 | 0.879 | 0.082-9.374 |
| | 16218 C>T | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 16223 C>T | 13.6 | 36.4 | 1.00 | 1.000 | 0.265-3.769 |
| | 16234 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16239 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16246 A>C | 2.3 | 6.8 | 1.00 | 0.879 | 0.082-9.374 |
| | 16260 C>T | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 16289 A>G | 4.5 | 4.5 | 0.29 | 3.000 | 0.372-24.171 |
| | 16290 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16298 T>C | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 16304 T>C | 11.4 | 29.5 | 1.00 | 1.044 | 0.271-4.015 |
| | 16311 T>C | 15.9 | 31.8 | 0.50 | 1.800 | 0.470-6.898 |
| | 16319 G>A | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16352 T>C | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16353 C>T | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 16362 T>C | 9.1 | 27.3 | 1.00 | 0.833 | 0.206-3.371 |
| | 73 A>G | 27.3 | 72.7 | - | - | - |

| | | | | | | |
|---|---|---|---|---|---|---|
| | 94 G>A | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 146 T>C | 6.8 | 27.3 | 0.50 | 0.556 | 0.125-2.465 |
| | 151 C>T | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 152 T>C | 15.9 | 25.0 | 0.18 | 2.673 | 0.686-10.412 |
| | 183 A>G | 4.5 | 2.3 | 0.17 | 6.200 | 0.507-75.838 |
| | 184 G>A | 4.5 | 2.3 | 0.17 | 6.200 | 0.507-75.838 |
| | 185 G>A | 6.8 | 4.5 | 0.11 | 5.000 | 0.720-34.726 |
| | 189 A>G | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 195 T>C | 4.5 | 6.8 | 0.60 | 1.933 | 0.281-13.295 |
| | 200 A>G | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 214 A>G | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 234 A>G | 2.3 | 4.5 | 1.00 | 1.364 | 0.112-16.577 |
| | 235 A>G | 2.3 | 2.3 | 0.47 | 2.818 | 0.162-49.008 |
| | 247 G>-A | 9.1 | 11.4 | 0.22 | 2.700 | 0.583-12.511 |
| | 263 A>G | 27.3 | 68.2 | 1.00 | - | - |
| | 302 A>AC | 9.1 | 20.5 | 0.72 | 1.278 | 0.307-5.320 |
| | 310 T>TC | 27.3 | 11.4 | 0.00 | - | - |

**Table 22: Association analysis between 66 variants found common in cases and controls.** * means the variants are associated with high risk factor.

Circos plot was used to represent the variants frequency distribution within BC cases. Even though the variants are specific to BC cases, we found that the variants observed are present in a relatively low frequency within the cases (Figure 22).

When the distribution of 24 common variants (cases and controls) were considered, 5 variants (C5178A, A15326G, C14766T, G13928C, G10653A) represented as C, U, S, R and O were equally distributed. The frequencies of 17 variants (G3316A, A4824G, A8860G, C9094T, A10398G, T10609C, G13708A, G15047A, T15449C and C15825T) were found in a much lesser frequency in cases than in the control groups. However, the frequency of two variants such as G12406A and C15402T were found to be higher in cases when compared to control groups (Figure 23).

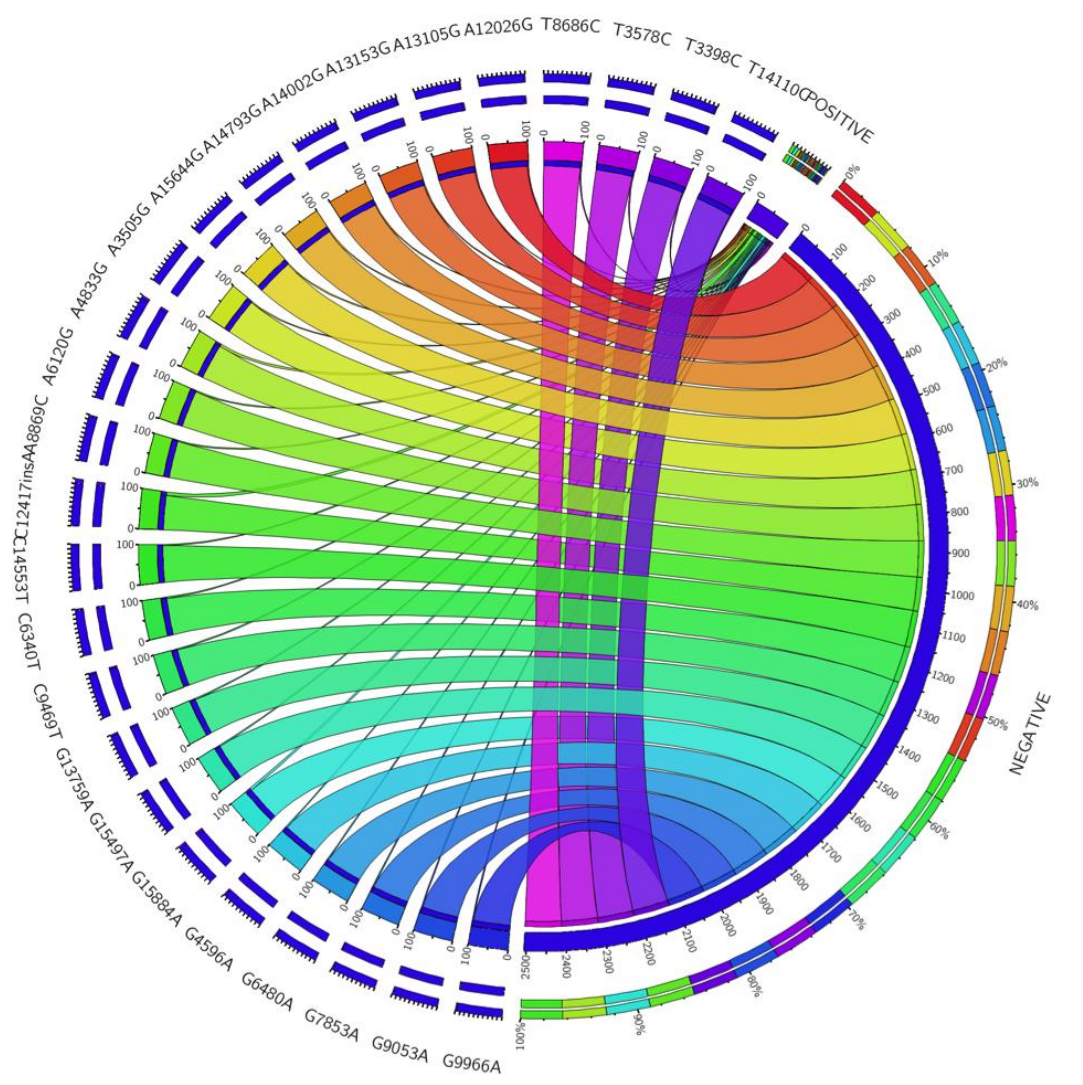



**Figure 22: Frequency distribution of case specific non-synonymous coding variants found in Breast cancer cases**. (a). List of variants in clockwise direction- G9966A, G9966A, G7853A, G6480A, G4596A, G15884A, G15497A, G13759A, C9469T, C6340T, C14553T, C12417insA, A8869C, A6120G, A4833G, A3505G, A15644G, A14793G, A14002G, A13153G, A13105G, A12026G, T8686C, T3578C, T3398C, T14110C. (b). Positive- Frequency representing presence of variants in BC samples (c). Negative- Frequency representing absence of variants in BC samples

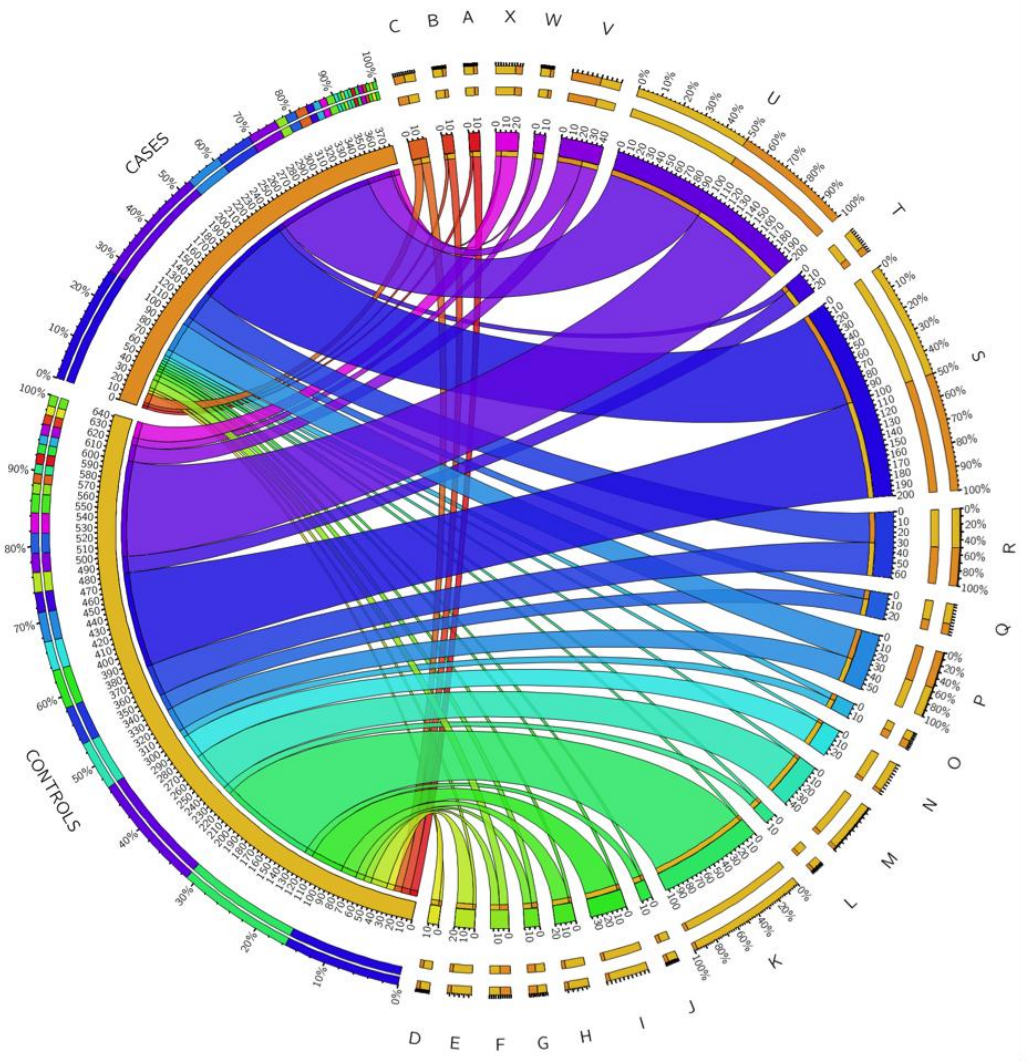



**Figure 23: Distribution of common variants found in cases and controls.**

(a). List of variants in alphabetical order:- A-G3316A, B- A4824G, C- C5178A, D-A5186T, E- G5460A, F- C8414T, G- G8584A, H- T8602C, I- A8701G, J- C8794T, K-A8860G, L- C9094T, M- A10398G, N-T10609C, O-G10653A, P- G12406A, Q-G13708A, R- G13928C, S- C14766T, T- G15047A, U- A15326G, V- C15402T, W-T15449C, X- C15825T. (b). Cases- BC samples. (c). Controls- Healthy control samples.

The BC cases resolved into two main clusters, when the variants were plotted against lifestyle risk factors. Most of the familial BC cases falls under one major cluster and was observed to have higher number of mitochondrial genes mutations. The major cluster showed familial BC and "tuibur" as the main risk factor. Dietary habits such as high consumption of smoked food, Sa-um and fish appear to be the major risk factor for BC. In case of non-familial BC cases, the frequency of mutations is elevated in mitochondrial respiratory complex 1 (ND1 – ND6) (Figure 22).

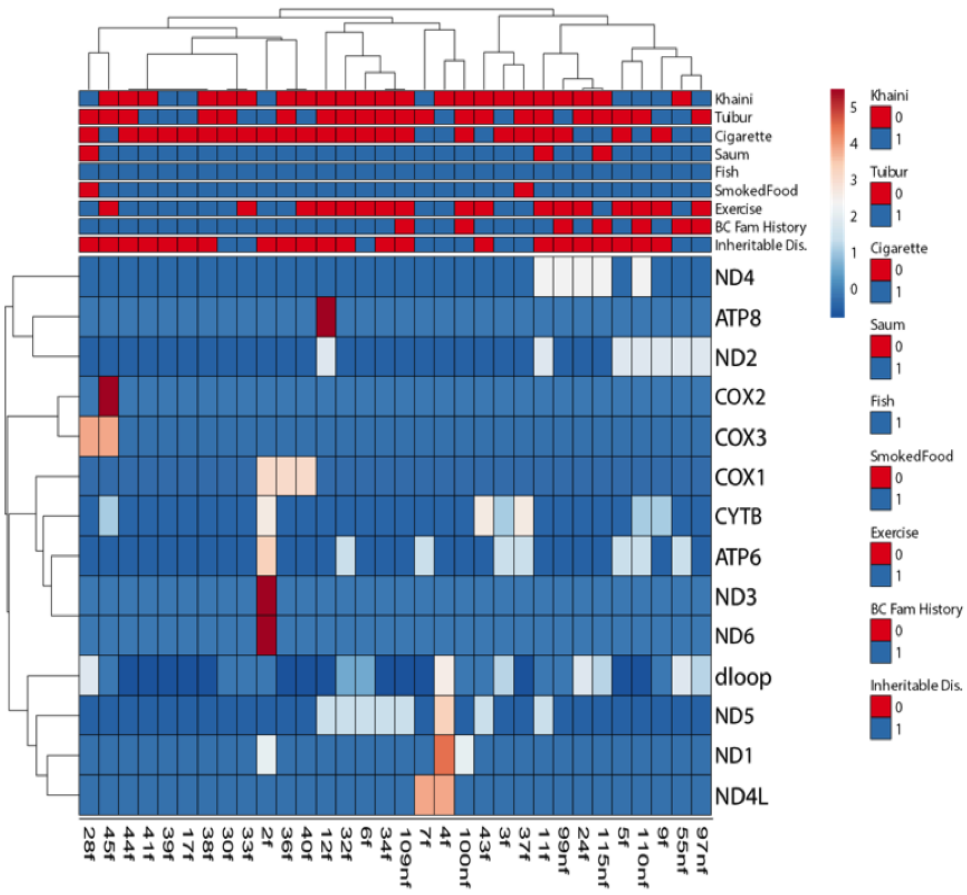



**Figure 24: Correlation of lifestyle habits with the frequency of variant in mitochondrial genome.**

# Chapter V

# Discussion

## ***Discussion***

This study focuses on assessing the possible role of several epidemiological factors and genetic alteration in BC progression using different approaches. BC is a multifactorial disease credited by certain genetics and epigenetic factors including reproductive factors, lifestyle and dietary habits (Sirisena et al., 2018). From the first analysis on the association between several epidemiological factors and the risk of BC, the incidence in Mizoram (maximum cases diagnosed at age 41- 50 years) was observed to be more or less similar to those observed in other Indian and foreign populations (Palachandra et al., 2017). This study on Mizo tribal population also revealed a similar and uniform dietary and life style habits across the samples (cases and controls). Interestingly, no distinctive identities for the major or minor tribes exist, since the people practice inter-marriages amongst themselves and no scientific studies have been carried out in this regard (The Wire, 2016). A positive association was observed between smoked food and sa-um with increased BC risk. Although there is a limited report on BC risk associated with smoked food, a 31% elevated all-cause mortality risk has been reported among BC patients with heavy intake of smoked food in their pre and post- diagnosis (Parada et al., 2017). Associations of dietary fats with BC risk have been supported by numerous case- control studies and animal experiments. Saturated fats in particular has been claimed as the main culprit for increasing endogenous estrogen production leading to an increase BC risk (Khodarahmi et al., 2014).

Khaini, a smokeless tobacco was found to increased BC risk. Many cancer incidence risk and smokeless tobacco usage has been reported (Accortt et al., 2005; Gupta et al., 2018). Among women with smokeless tobacco users living in the tribal lands of Western North Carolina, an 8 fold increased risk of BC has also been reported (Spangler et al., 2000; Science Daily, 2000). Lower amount of vegetables and water intake act as a confounding factors for BC progression. A higher chance of BC development has been reported for women consuming less vegetables in comparison to those with higher intake. Vegetables provide several anti- cancer drugs, phytochemicals and protease inhibitors (Kooshki et al., 2016). The risk of BC could be reduced in pre-menopausal and post-menopausal women by 33% and 79%, respectively upon hydration. Cancer may arise from failure of harmful toxin removal due to impairment of internal function caused by dehydration (Stookey et al., 1997).

Our observation on the potential BC risk associated with having a 1$^{st}$ and/ or 2$^{nd}$ degree relative with BC has been abundantly supported by 38 studies that suggested the relative risk of developing BC is 2.1 for those with 1$^{st}$ degree relative with BC (Mahdavi et al., 2019). Studies on UK cohort also report a 2.5 fold increase risk of BC among women with more than two relatives diagnosed with BC (Brewer et al., 2017). Diabetes and hypertension were frequently observed within the family of BC patients and was reported to have associated with an increase risk of BC development. Available records from different population on diabetes associated BC risk has been very contrasting (Maskarinec et al., 2018; Hardefeldt et al., 2012). However, the inconsistency in results may be attributed to the adipokines and visceral fats difference of different ethnicity, quantity of glucose and chronic

inflammation governed by the metabolic activity of obesity, altering the association between BC and diabetes (Maskarinec et al., 2018).

Among postmenopausal hypertensive Asian women, hypertension has been found to increase BC risk by 15%. Hypertension and BC has been assumed to share a pathophysiological pathway facilitated by adipose tissues causing chronic inflammation, further leading to an elevated risk of both the diseases. Furthermore, alteration of apoptosis by hypertension might further affect the regulation of cell turnover, thereby increasing BC risk (Han et al., 2017). A well-established BC risk factors for different population is late age at menopause, which is also observed in our studies. Elongated and elevated duration of ovarian hormone exposure may be the probable reasons (Khalis et al., 2018).

Shorter duration of lactation is also found to be the major risk factors for BC. Breastfeeding plays a protective role against BC by lessening the intensity and ovulation frequency, which in turn lowers the level of estrogen (Lodha et al., 2011). Extended duration of lactation has also been shown to reduce BC risk among women with BRCA1 gene mutation (Freund et al., 2005). As reported by the World Cancer Research/ American Institute for Cancer Research in 2007, there is an inverse relation between physical activity and postmenopausal BC risk (Hildebrand et al. 2013). Lack of physical activity was found to be a confounding factor for BC in Mizo population. The physical activities role in protection against BC risk has been reported by many studies and has been attributed in modifying the levels of sex hormone, adiposity, immune function, and insulin-related hormones (Loprinzi et al., 2012).

From the univariate analysis, since fish was found to have an independent association with increase risk of BC. In Mizoram, fish becomes an essential part of diet since it is the main source of protein and omega-3 fatty acids (Environmental Defense Fund, 2019). However, Mizoram being a hilly place has many disadvantages in fish cultivation. This brings the demand of importing fish from neighboring states and border countries. Fishes in Mizoram were mainly obtained from Andhra Pradesh, Burma and Silchar. Fish perishable were customarily injected or dipped in a formalin solution to make it firm, enhanced shelf life, and to make it look fresh for longer period of time (Uddin et al., 2011). Formalin estimation was carried out on fish samples from 4 different sources and was found that fish from Andhra Pradesh has the highest concentration of formalin (3.17 µg/g), followed by Silchar (1.52 µg/g), Burma (0.28 µg/g) and Mizoram (0.17 µg/g). According to the United States Environmental Protection Agency (EPA), formaldehyde used should not exceed more than 0.2 µg/g body weight per day (Noordiana et al., 2011). Majority of the fish available is from Andhra Pradesh. However, fish from Andhra Pradesh have far exceeded the recommended concentrations of formalin. International Agency for Research on Cancer has grouped formalin as Group 1 carcinogenic to humans. (IARC, 2004)

Second, the association of histological tumor grade with other clinical and epidemiological features was analyzed. Histological tumor grade serves as the best prognostic factor for BC, since it can deliver fundamental reports about BC clinical behavior by assessing the morphology of tumor biology (Rakha et al., 2010). We observed an association between tumor grade with lymph node status in this study. Lymph node has also been considered as a critical BC prognostic factors, since it can help determine cancer stage and treatment option. Rate of survival decreases and recurrence rate increases as the number of positive axillary lymph nodes increases (Siadati et al., 2015; Davis et al., 1986). Better survival has been observed in lymph node negative patient when compared to the positive patient. Additionally, regardless of the tumor size, poorest survival outcome has been observed among patient with more than 10 positive lymph nodes (Wang et al., 2016).

Our result on the association of tumor grade with ER has been supported by the findings on Western Chinese populations (Zheng et al., 2018). Decrease in ER+ was observed as the tumor grade increases. Similar with the study performed by Atif et al. (2018), tumor grades I and II have a higher number of ER+ when compared to grade III tumor. Elevated expression of an epidermal growth factor called HER2/neu oncoprotein found on the cell surfaces and transmit growth signal to nucleus is associated with poor prognosis. We observed significant correlation between tumor grades with HER-2/neu oncoprotein. Several studies have reported a positive correlation between high grade tumor and HER-2/neu expression (Siadati et al., 2015; Zheng et al., 2018). In our study, we observed an inverse relationship between the tumor grade with hormone receptors (ER and PR). As the tumor grade increases,

the number of ER+/PR+ was found to be decreasing. Our result is in corroboration with several other results (Atif et al., 2018; Effi et al., 2017; Zheng et al., 2018). This result can help provide treatment options such as the ER+/PR+ patients for hormonal treatment, while ER- PR- patients for chemotherapy.

Triple negative BC subtype is well known characterized by early onset of BC, poor prognosis, extreme invasiveness, early recurrence (within the first 3 years) and high 5-year mortality rate (Wang et al., 2016; DeGennaro Jr et al., 2018). The prevalence of this triple subtype is found to differ among different population. India has been reported to have considerably higher rate of triple negative BC (31%) when compared to Western populations (Sandhu et al., 2016). The prevalence of triple negative subtype in our study is 17.47%, which fall in line with the findings on Southeastern Turkey population (Mirunalini et al., 2010; Sandhu et al., 2016). We also found a significant association between tumor grade with status of ER/PR/HER2. Several studies had performed correlation, separately between tumor grade with ER/PR and then with HER2/neu status. A significant correlation between tumor grade I and II with ER+/PR+ and tumor grade with HER2/neu oncoprotein has been consistently reported (Siadati et al. 2015).

In our study, having a first degree relative with BC is found to be significantly associated with tumor grade. One of the most prominent risk factor for BC onset is family history with the associated relative risk accounted by number of individuals affected in the family, their age at disease onset and their degree of relationship (Tazzite et al. 2013). Previous correlation analysis results between tumor

grade with family history of BC has been contradicting. Melvin et al. (2016) performed association study between family history of BC and severity of the disease at the time of diagnosis and found no association, neither with severity or mortality. However, another study have reported a higher risk of poorer survival (relative risk-1.54, 95% CI: 0.98–2.41) for BC patient, age less than 50 years and having a first degree relatives diagnosed with BC (Slattery et al. 1993).

The first approach that was carried out to assess the prevalence of genetic susceptibility in Mizo BC patients was screening of candidate genes such as *BRCA1, TP53, PTEN, CDH1, CHEK2* and *XRCC2.* From this analysis, eight polymorphisms were observed in BRCA1 gene. Even though the polymorphisms presented both homozygous and heterozygous conditions, there was no significant amino acid change observed. They were previously reported by various genetics clinics to have no clinical value in databases like BIC BRCA ((https://research.nhgri.nih.gov/bic/) and ClinVar (https://www.ncbi.nlm.nih.gov/clinvar/) and most of them are unpublished (Sharma-Oates et al., 2018). Although, the exons for other candidate genes were chosen based on more number of codons reported for BC, we do not observed any genetic alteration, which could be because of sample size.

p.S694S has been described to have association with age at diagnosis (Ricks-Santi et al., 2017). These polymorphisms were inherited through shared haplotype that were in significant linkage disequilibrium (LD). Polymorphisms such as p.L771L, p.S1436S, p.P871L, p.E1038G and p.K1183R were also found in Indian population and among them, p.P871L, p.E1038G and p.K1183R were included in top

20 BIC entries. Presence of both p.K1183R and p.P871L has been reported to decrease the risk of BC and overall cancer, respectively, proposing their protective role (Buleje et al., 2017; Sharma-Oates et al., 2018; Miao et al. 2017). Incidence of cancer despite the presence of p.K1183R and p.P871L might be explained by the individual's susceptibility to BC that is determined by the equilibrium between DNA damage and repair (Miao et al. 2017). Though synonymous substitution, p.P1544P observed in exon 15 was found to be novel. Surprisingly, only two genetics clinics has reported p.D1546Y of exon 15 for European (Non Finnish) population with MAF>0.00003295 in BIC BRCA (https://research.nhgri.nih.gov/projects/bic/Member/cgi-bin/bic_query_result.cgi?table=BRCA1_exons&nt=4755&base_change=G%20to%20T&exact_search=1) and ExAC http://exac.broadinstitute.org/variant/17-41226387-C-A).

As a second approach, targeted re-sequencing of BRCA1 and BRCA2 genes was performed for 15 samples (10 BC cases and 5 healthy controls). Using this approach, two pathogenic mutations were identified in BRCA1 gene. They are p.Gln126Ter (p.Q126X) in exon 7 and p.Cys1697Arg (p.C1697R) in exon 18. The substitution of glutamine into a premature stop codon (CAA>TAA) generates a nonsense mutation that interrupts the function of standard normal protein. It is predicted that the proteins lose their normal function due to protein truncation generated by occurrence/insertion of premature stop codon or nonsense-mediated mRNA decay. This kind of mutations that leads to protein truncation are known to be functionally deleterious. Even though few genetic clinics has reported this pathogenic mutations in ClinVar, dbSNP and 1000 Genomes, to our knowledge, no

existing literature has been found reporting this mutation (https://www.ncbi.nlm.nih.gov/clinvar/variation/427022/; Vallon-Christersson et al., 2001).

The next mutation, p.Cys1697Arg was reported thrice in BIC database (https://research.nhgri.nih.gov/projects/bic/Member/cgi-bin/bic_query_result.cgi?exact_search=1&table=brca1_exons&exon_type=18), once by a genetic clinic and the others from a Danish population. It is frequently identified among the Danish Hereditary Breast and Ovarian Cancer (HBOC) families (Thomassen et al., 2008). The substitution of cysteine to arginine is rather dramatic because the wild type residue is non-polar amino acid, having the ability of forming disulfide linkages whereas the mutant residue has a positive charge found in an undesirable α-helix based on the structure of XRCC1 BRCT (BRCA1 C-terminal) (Zhang et al., 1998). This alteration is identified in the BRCT domain of BRCA1 gene, where a stretch of 23 amino acids is conserved among human, mouse and canine (Szabo et al., 1996). To further support our finding, functional analysis of C1697R for loss of C-terminal transactivation activity suggested that activation of transcription was not possible with BRCA1 arg1697 construct (Vallon-Christersson et al., 2001). In this study, C1697R was identified in 38 year old patient having a family history of BC, Ovarian cancer and other cancer types in her first and second degree relative. The same mutation was identified in 46 year old Danish BC patient having the same condition of family history (Bergthorsson et al., 2001). Additionally, this mutation was found in only a particular family when screening was done on 450 BC cases having familial history of breast and ovarian cancer, which indicates that C1697R characterize rare

variants (Vallon-Christersson et al., 2001). Pathogenic variants were not observed in BRCA2 gene in the present study samples. It is possible that other BC predisposing gene might play a role in the development BC progression, rather than BRCA2 gene (Frey et al., 2017).

Clinical exome panel sequencing was performed for twelve samples, consisting of seven familial BC cases, two non-familial cases and three healthy controls. We observed twenty-eight variants, in which five variants were found repeating in both the approaches used for analysis. The variants were all classified based on the recommendation of American College of Medical Genetics and Genomics- Association for Molecular Pathology (ACMG-AMP) guidelines such as benign, likely benign, variants of uncertain significance (VUS), pathogenic and likely pathogenic, which have been widely used in variant classification and also using the software prediction tools (Li et al., 2018; Richards et al., 2015; Hampel et al., 2015).

Using the first approach (that excluded all variants belonging to control groups, synonymous variants and minor allele frequency > 0.05), 16 variants were observed which represented 3 novel, 5 purely benign, 1 likely benign and 7 VUS. These novel variants belong to BLM, CDKN2A and FANCC susceptibility genes confers increased risk for BC (Rossing et al., 2019). From the protein structural prediction of novel missense variants such as BLM c.485C>T: p.Thr162Ile and FANCC c.590A>G: p.Asp197Gly, the 3D-structure or modeling template could not be obtained for both the substitution. For BLM c.485C>T: p.Thr162Ile, both the

wild-type and mutant residues differ in properties such as size, charge and hydrophobicity value. The mutant residue, isoleucine is much bigger than wildtype threonine, which might lead to collisions. Also since isoleucine is more hydrophobic, introduction of more hydrophobic residue might lead to reduction of hydrogen bonds, which might further hamper correct protein folding. Threonine is found to be very conserved, mutation in this position is predicted to be probably damaging to the protein (Venselaar et al., 2010).

For FANCC gene, c.590A>G: p.Asp197Gly, Glycine, the mutant residue, is smaller and neutral in charge while the wild-type, aspartate is much bigger and negative in charge, which might disturb the interactions with other residues or molecules. Glycines are also very flexible and might also disrupt the required firmness of the protein at this position. Hydrophobicity is more in the mutant residue than in the wild-type residue and consequences of introducing a more hydrophobic amino acid at this location is loss of hydrogen bonds and/or intrude with proper protein folding (Venselaar et al., 2010). Eventhough, functional validation has not been performed for these novel variants, they might play a vital role for BC progression in Mizo population.

Certain VUS observed such as, CDH1 p.Glu880Lys, GEN1 p.Ile203Val and p.Ser873Arg and MUTYH p.Ala370Val has been reported in other populations. Eventhough, their clinical significance has not been reported, it cannot be neglected, since several important variants may be misclassified as VUS due to their extreme rarity and/or heterogeneity. Increasing the study sample size alone cannot

significantly promote the efficiency of BC predisposition gene identification (Kohane et al., 2006).

By considering the different algorithm of mutation prediction tools, several missense, frameshift and intronic splice site variants belonging to ATM, BRCA2, CDH1, CDKN2A, MRE11A, MSH6, MUTYH, PALB2, STK11 and TP53 were identified among the Mizo BC patients and has already been reported. Among these variants, structural prediction was performed for p.Gly283Glu in MUTYH and p.Arg181His in TP53 gene since they were further predicted to be pathogenic according to ACMG guidelines. According to this prediction, the substitution of Glycine by Glutamine at position 182 is probably damaging since it is located in an important domain essential for the protein activity and also a region of interaction with other residues of another domain. Furthermore, glycine being flexible and smaller in size is found buried inside the protein structure with an unusual torsion angles. Hence, glutamine being bigger will not fit in this place and will force a fault conformation of the local backbone, which will further disturb the protein structure (Venselaar et al., 2010). BRCA1 and BRCA2 gene variations could explain only about 15% of increased risk for BC development, the remaining risk could be explained by mutation in several other moderate and low penetrance genes such as ATM, BARD1, BLM, STK11, NF1, BRIP1, CDH1, CHEK2, FANCC, RAD51, TP53, PTEN, PALB2, PMS2 (Easton et al., 2015). Therefore, in Mizo BC population, it is not surprising to identify sequence variations in diverse panel genes other than BRCA1 and BRCA2 genes. However, technical, structural and functional validation using other approaches such as Sequenom MassARRAY platform

(Sequenom, San Diego, CA) (Zhang et al., 2015) or direct Sanger sequencing is required to confirm the prevalence or contribution of these variants for BC in Mizo population.

The last approach was screening of mtDNA gene mutation and their role in breast carcinogenesis. One novel frameshift (12014C>CA; c.1255_1255delinsCA) variant in ND5 gene was identified in the Mizo population that might act as a potential risk factor for BC. Having no structural impact on protein, ATP6 8860A>G was found as common variants in our study, and even though it is found in less conserved region it might still be responsible for other mitochondrial and nuclear DNA mutations. This variant has been reported in 100% of BC tumor samples and has been suggested to increase BC risk significantly ($p < 0.05$) (Ghaffarpour et al., 2014). Additionally, patients with more non-synonymous variation in ATPase gene had been suggested to have a shorter survival when compared to those with only one variation (Zhu et al., 2005). A10398G found in ND3 gene from the samples had been a variant of interest as a possible candidate marker for BC development among the American women ((Czarnecka et al., 2010). While the allele 10398G increased BC risk among the European- American (Bai et al., 2007), Polish (Czarnecka et al., 2010) and Malay (Nadiah et al., 2012) women, allele 10398A is a risk factor for North Indian (Darvishi et al., 2007) and African- American (Canter et al., 2005) women, further highlighting the important role of haplotype difference due to mtDNA polymorphisms. This substitution leading to changes in amino acid threonine to alanine (Bai et al., 2007) has been referred as a reliable biomarker and an independent signal for BC development (Weigl et al., 2013).

In the present study, several variants that were found to be affecting the normal protein function (based on mutation prediction tools) and also those found in the non-coding regions (reported for numerous other diseases) might contribute to BC susceptibility in the Mizo population. Triggered by the environmental factors, it is well known that mtDNA polymorphisms or different alleles of the same locus could act as disease risk factors in certain haplogroup (Canter et al., 2005). We observed T16126C in hormone receptor positive BC. In contrast, a higher frequency of this variant was found in hormone receptor negative BC (47%) when compared to hormone receptor positive BC (12%) (Tommasi et al. 2014). Variants such as A73G, T152C, A263G, T16189C, C16223T, C16290T, T16304C, T16362C found in the d-loop region has been found reported for BC in many populations (Czarnecka et al., 2010; Cai et al., 2011; Tommasi et al. 2014). Although, previous studies had suggested A73G, C16223T and T16362C to have a protective effect against BC, since they were found at a higher frequency in control groups, we observed a higher frequency of these variants in cases suggesting that they might contribute to BC occurrence in this population (Zhu et al., 2005). Other studies have frequently reported T152C, A263G and T16189C for their association with BC risk (Jiménez-Morales et al., 2018).

Non-synonymous substitutions in both the mitochondrial respiratory complex I and IV (dN/dS ratio of 3.24 and 1.71, respectively) were found to be positively selected suggesting their driver role in BC. Since each of these complexes functions as an important unit, so non-synonymous substitutions in these complexes may disturb the normal purpose of those particular complexes (Palodhi et al., 2018).

High amount of non- synonymous substitution and less synonymous substitution in the protein coding genes indicates a positive selection during tumorigenesis and such mutational pattern may represent the existence of genuine driver mutations in certain cancers (Hodis et al., 2012). Furthermore, study had shown that enrichment of non-synonymous substitution of mitochondrial gene in BC tumor samples taken from TCGA data and also positive selection of those mutations (McMahon et al., 2014). Two high-risk variants (8701 A>G and 10609 T>C) resulted in missense and frameshift substitution leads to enhanced mutation in nuclear gene, thereby disrupting apoptosis that plays a vital role in the formation of BC (Li et al., 2016).

# Chapter VI

# Summary and Conclusion

## ***Summary and Conclusion***

Breast cancer is the leading cancer site and major cause of cancer death in women and has become the global burden. This complex multifactorial disease regulated by genetic, hormonal and epigenetic factors such as diet and lifestyle habits has been an area of research interest. Decades of scientific research have lead us to the understanding of a fraction of the causal factors such as mutations in susceptibility nuclear as well as mitochondrial genes and also environmental factors which differs in different ethnic groups. Despite the groundbreaking advances being made in term of the discovery of more genes involved using high- throughput sequencing technologies and improve strategies for early detection and treatment options, there is a consistent increase in the prevalence of the disease.

The present study on Characterization of clinically significant mutations associated with Breast Cancer in Mizo population has been carried out with the aim of evaluating the demographic risk factors and assess the prevalence of mutation in susceptibility genes associated with breast cancer in Mizo population, Mizoram.

The findings of the present work are summarized as follows:

- In Mizoram, BC cases is observed from age as early as 20 and as old as 91 years. The mean average age was observed to be 49 years.
- Environmental factors such as inheritable diseases, having 1st and 2nd degree relative with BC, high consumption of khaini (tobacco), smoked food, sa- um (fermented pork fats), lack of physical exercises and shorter lactation duration were found to be the significant risk factors for BC in Mizo population.

- Other factors including late age at menopause, lesser intake of water and vegetables were also found to act as a potential confounding factors among the Mizo BC patient.
- We have observed that with the increase in the BC tumor grade, there is more of positive lymph node invasion. Tumor grade was found to have a significant association with lymph node invasion ($p<0.021$), ER ($p< 0.004$) and Her2/neu ($p< 0.014$) independently, double subtype of ER/PR ($p< 0.007$) and also the triple subtype ER/PR/HER2 ($p< 0.025$).
- The study showed that having at least one or more first degree relative's BC have a significant association with tumor grade ($p<0.003$).
- Interestingly, other variables such as tumor characteristics, reproductive factors, dietary and lifestyle habits does not show any significant association with tumor grade.
- Eight polymorphisms such as p.L771L, p.S694S, p.P871L, p.E1038G, p.K1183R, p.S1436S, p.P1544P, p.D1546Y were observed in BRCA1 gene from the candidate gene analysis using Sanger sequencing.
- Among them, p.P1544P in exon 15 was found to be novel. There was no significant amino acid change to be noted since all the polymorphisms represent silent substitutions.
- From targeted re-sequencing of BRCA1 and BRCA2 gene, two pathogenic mutations such as p.C1697R and p.Q126X in BRCA1 gene were observed in two familial Mizo BC samples. Among the familial BC samples included in the targeted re- sequencing, no pathogenic variants were observed in BRCA2 gene.

- Using clinical exome sequencing, the analysis based on ACMG (American College of Medical Genetics) variants classification guidelines, 16 variants consisting of 5- benign variants, 3- either benign or VUS (Variants of Uncertain Significance), 1- likely benign and 7 VUS were observed.
- Among them, 3 variants in BLM (p.Thr162Ile), CDKN2A (intronic) and FANCC (p.Asp197Gly) were found to be novel.
- The second approach for the analysis of clinical exome sequence data was performed using the software prediction tools such as SIFT, Polyphen2, Provean, LRT, FATHMM, METASVM, METALR and CADD. Twelve (12) deleterious variants in the Mizo population were observed: p.Arg2060Cys, p.Asn289His, p.Glu880Lys, p.Gly62Ser, CDKN2A (Intronic-ss-acr), p.Arg87Trp, p.Lys1358AspfsTer2, p.Ala370Val, p.Gly283Glu, p.Asp498Tyr, p.Thr363Ile, p.Arg181His.
- From whole mitochondrial sequencing, sequence alterations were found to be the highest in d-loop region and least in ND6 region.
- In the protein coding region, 51 variants were observed, in which 27 were BC specific and 24 were found in common among the cases and controls.
- One frameshift variants (12014C>CA; c.1255_1255delinsCA) located in ND4 gene was found to be novel. Four variants (ATP6 8860A>G, ND3 10398A>G, ND5 13708G>A and ND5 14110T>C) found common in both cases and controls have been reported in database for BC.
- According to SIFT and Polyphen- 2 software analysis, 22 variants were found to be affecting the normal protein functions. Based on SOSUI software, 98% of non- synonymous variants falls in the hydrophobic region.

- The potential impact of a greater number of non-synonymous substitutions in the protein coding region represented decrease in protein stability, whereas, only two variant in ND2 and ND6 with ΔΔG of 0.144 and 0.025, respectively, were observed with increase protein stability according to Pmut.
- In the non-coding d-loop region of mitochondrial genome, 68 variants comprising of 26 BC specific variants and 42 common variants was observed.
- The variant specific to BC cases, 16126T>C and other eight common variants such as 73A>G, 152T>C, 263A>G, 16189T>C, 16223C>T, 16290C>T, 16304T>C, 16362T>C has been reported for BC.
- The dN/dS ratio of the variants in complex I and III shows a positive selection for non- synonymous variants indicating the driver role of these complexes in BC development.
- According to correlation analysis, ATP6 8701A>G, ND3 10398A>G, ND4L 10609T>C in coding region and 16188 C>-T in the d-loop region were found to be high risk factors for BC occurrence in Mizo population.
- Analysis of variant clustering on mitochondrial genome based on the lifestyle risk factors showed that most of the familial BC cases falls under one major cluster and is observed to have great number of mitochondrial gene mutations.
- The major cluster showed familial BC and “tuibur” as the main risk factor. Dietary habits such as, high consumption of smoked food, Sa-um and fish are the major risk factor for BC.
- In case of non-familial BC cases, the frequency of mutations is elevated in mitochondrial respiratory complex 1 (ND1 – ND6).

The present study is the first scientific evaluation of the demographic, epidemiological and environmental risk factors towards BC progression and assessment of BC panel gene mutations in the population of Mizoram. The outcome of this study on several epidemiological risk factors can serve as baseline data to provide a more confined risk factors in different age groups of the population. Furthermore, mutation studies provide us with the knowledge on the prevalence of several BC susceptibility nuclear as well as mitochondrial gene mutations. Thus, from this study we can conclude that even though several reproductive, dietary and lifestyle factors play an important role in the development of BC, the extent of the contribution of several genes mutations need further validation in order to fully understand their role in BC progression in Mizo population.

## References

Accortt NA, Waterbor JW, Beall C, Howard G (2005) Cancer incidence among a cohort of smokeless tobacco users (United States). Cancer Causes and Control 16:1107-1115.

Agarwal G, Ramakant P (2008) Breast cancer care in India: The current scenario and the challenges for the future. Breast Cancer 3:21-7.

Alghamdi IG, Hussain II, Alghamdi MS, El- Sheemy MA (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.

Anstey EH, Shoemaker ML, Barrera CM, O'Neil ME, Verma AB, Holman DM (2017) Breastfeeding and Breast Cancer Risk Reduction: Implications for Black Mothers. American Journal of Preventive Medicine 53(3S1):S40-S46.

Apostolou P, Fostira F (2013) Hereditary breast cancer: the era of new susceptibility genes. BioMed research Int 747318 doi:10.1155/2013/747318.

Atif N, Khalid M, Chughtai O, Asif Saad, Rashid M, AS Chughtai (2018) Role of immunohistochemical markers in breast cancer and their correlation with grade of tumour, our experience. International Clinical Pathology Journal 6(3):141-145.

Bai R, Leal SM, Covarrubias D, Liu A, Wong LC (2007) Mitochondrial Genetic Background Modifies Breast Cancer Risk. Cancer Research 67:(10).

Bergthorsson JT, Ejlertsen B, Olsen JH, Borg A, Nielsen KV, Barkardottir RB, Klausen S, Mouridsen HT et al., (2001) BRCA1 and BRCA2 mutation status and cancer family history of Danish women infected with multifocal or bilateral breast cancer at a young age. Journal of Medical Genetics 38:361–368.

Bianchi MS, Bianchi NO, Bailliet G (1995) Mitochondrial, DNA mutations in normal and tumor tissues from breast cancer patients. Cytogenetics and Cell Genetics 71:99–103.

Brewer HR, Jones ME, Schoemaker MJ, Ashworth A, Swerdlow AJ (2017) Family history and risk of breast cancer: an analysis accounting for family structure. Breast Cancer Research and Treatment 165(1):193-200.

Buleje J, Guevara-Fujita M, Acosta O, Huaman FDP, Danos P, Murillo A, Pinto JA, Araujo JM et al., (2017) Mutational analysis of BRCA1 and BRCA2 genes in Peruvian families with hereditary breast and ovarian cancer. Molecular Genetics & Genomic Medicine 5(5):481–494.

Cai FF, Kohler C, Zhang B, Chen WJ, Barekati Z, Garritsen HSP, Lenner P, Toniolo P, Zhang JJ, Zhong XY (2011) Mutations of Mitochondrial DNA as Potential Biomarkers in Breast Cancer. Anticancer Research 31:4267-4272.

Campeau PM, Foulkes WD, Tischkowitz MD (2008) Hereditary breast cancer: new genetic developments, new therapeutic avenues. Human Genetics 124: 31-42.

Canter JA, Kallianpur AR, Parl FF, Millikan RC (2005) Mitochondrial DNA G10398A Polymorphism and Invasive Breast Cancer in African-American Women. Cancer Research (17)65.

Crawford B, Adams SB, Sittler T, Akker JVD, Chan S, Leitner O, Ryan L, Gil E et al., (2017) Multi-gene panel testing for hereditary cancer predisposition in unsolved high-risk breast and ovarian cancer patients, Breast Cancer Research and Treatment 163:383–390.

Czarnecka AM, Klemba A, Krawczyk T, Zdrozny M, Arnold RS, Bartnik E, Petros JA (2010) Mitochondrial NADH-dehydrogenase polymorphisms as sporadic breast cancer risk factor. Oncology Reports 23:531-535.

D'Argenio V, Esposito MV, Telese A, Precone V, Starnone F, Nunziato M, Cantiello P, Iorio M et al., (2015) The molecular analysis of BRCA1 and BRCA2: Next-generation sequencing supersedes conventional approaches. Clinica Chimica Acta 446:221–225.

Darvishi K, Sharma S, Bhat AK, Rai E, Bamezai RNK (2007) Mitochondrial DNA G10398A polymorphism imparts maternal Haplogroup N a risk for breast and esophageal cancer. Cancer Letters 249:249–255.

Davis BW, Gelber RD, Goldhirsch A, Hartmann WH, Locher GW, Reed R, Golouh R, Save-Doderbergh J et al., (1986) Prognostic Significance of Tumor Grade In Clinical Trials of Adjuvant Therapy for Breast Cancer With Axillary Lymph Node Metastasis. Cancer 58:2662-2670.

DeGennaro Jr V, Jiwani F, Patberg E, Gibbs M, Libby R, Gabriel D, Heldermon CD, Daily K, Jr JB (2018) Epidemiological, Clinical, and Histopathological Features of Breast Cancer in Haiti. Journal of Global Oncology doi.org/10.1200/JGO.17.00135.

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. (2011) A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nature Genetics 43(5):491–8.

Zodinpuii D, Pautu JL, Zothankima B, Pachuau L, Kumar NS (2019) 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 doi.org/10.1007/s11356-019-06944-8.

Easton DF, Pharoah PD, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, et al., (2015) Gene-panel sequencing and the prediction of breast-cancer risk. New England Journal of Medicine 372:2243–57.

Easton DF, Pharoah PDP, Antoniou AC, Tischkowitz M, Tavtigian SV, Nathanson KL, Devilee P, Meindl A et al., (2015) Gene-Panel Sequencing and the Prediction of Breast-Cancer Risk. New England Journal of Medicine 372;23.

Effi AB, Aman NA, Koui BS, Koffi KD, Traore ZC, Kouyate M (2017) Immunohistochemical determination of estrogen and progesterone receptors in breast cancer: relationship with clinicopathologic factors in 302 patients in Ivory Coast. BMC Cancer 17:115.

Freund C, Mirabel L, Annane K, Mathelin C (2005) Breastfeeding and breast cancer. Gynecology Obstetrique anf Fertilite 33:739-44.

Frey JD, Salibian AA, Schnabel FR, Choi M, Karp NS (2017) Non-BRCA1/2 Breast Cancer Susceptibility Genes: A New Frontier with Clinical Consequences for Plastic Surgeons. Plastic and Reconstructive Surgery Global Open 5:1564.

Gage M, Wattendorf D, Henry LR (2012) Translational advances regarding hereditary breast cancer syndromes. Journal of Surgical Oncology 105(5):444–451.

García-Alcalde F, Okonechnikov K, Carbonell J, Cruz LM, Go□tz S, Tarazona S, Dopazo J, Meyer TF, Conesa A (2012). Qualimap: evaluating next-generation sequencing alignment data. Bioinformatics 28(20):2678-9.

Ghaffarpour M, Mahdian R, Fereidooni F, Kamalidehghan B, Moazami N, Houshmand M (2014) The mitochondrial ATPase6 gene is more susceptible to mutation than the ATPase8 gene in breast cancer patients. Cancer Cell International 14:21.

Gupta S, Chaurasia A, Pathak E, Mishra R, Chaudhry VN, Prashaant Chaudhry, MDd, Ashim Mukherjee, PhDa, Mousumi Mutsuddi (2017) Whole exome sequencing unveils a frameshift mutation in CNGB3 for cone dystrophy. A case report of an Indian family. Medicine 96:30(e7490) doi.org/10.1097/MD.0000000000007490

Gupta S, Gupta R, Sinha DN, Mehrotra R (2018) Relationship between type of smokeless tobacco & risk of cancer: A systematic review. Indian Journal of Medical Research 148(1):56-76.

Hampel H, Bennett RL, Buchanan A, Pearlman R, Wiesner GL (2015) A practice guideline from the american college of medical genetics and genomics and the national society of genetic counselors: referral indications for cancer predisposition assessment. Genetics in Medicine 17:70–87.

Han H, Guo W, Shi W, Yu Y, Zhang Y, Ye X, He J (2017) Hypertension and breast cancer risk: a systematic review and meta- analysis. Scientific Reports 7:44877.

Haraldsdottir A, Torfadottir JE, Valdimarsdottir UA, Adami HO, Aspelund T, Tryggvadottir L, Thordardottir M et al., (2018) Dietary habits in adolescence and midlife and risk of breast cancer in older women. PLoS ONE 13(5):e0198017.

Hardefeldt PJ, Edirimanne S, Eslick GD (2012) Diabetes increases the risk of breast cancer: a meta-analysis. Endocrine-Related Cancer 19 793–803.

He C, Chasman DI, Dreyfus J, Hwang SJ, Ruiter R, Sanna S, Buring JE, Fernández-Rhodes L et al., (2012) Reproductive aging-associated common genetic variants and the risk of breast cancer. Breast Cancer Research 14:R54

Hildebrand JS, Gapstur SM, Campbell PT, Gaudet MM, Patel AV (2013) Recreational physical activity and leisure-time sitting in relation to postmenopausal breast cancer risk. Cancer Epidemiology, Biomarkers and Prevention 22(10):1906–12.

Hodis E, Watson IR, Kryukov GV, Arold ST, Imielinski M, Theurillat JP, Nickerson E, Auclair D et al., (2012) A Landscape of Driver Mutations in Melanoma. Cell 150(2):251–263.

International Agency for Research on Cancer. 2004. Monographs on the Evaluation of Carcinogenic Risks to Humans, Vol. 88, International Agency for Research on Cancer, Lyon, France.

Jaman N, Hoque MS, Chakraborty SC, Hoq ME, Seal HP (2015) Determination of formaldehyde content by spectrophotometric method in some fresh water and marine fishes of Bangladesh. International Journal of Fisheries and Aquatic Studies 2(6): 94-98.

Jian W, Shao K, Qin Q, Wang X, Song S, Wang X (2017) Clinical and genetic characterization of hereditary breast cancer in a Chinese population. Hereditary Cancer in Clinical Practice 15:19 DOI 10.1186/s13053-017-0079-4.

Jiménez-Morales S, Pérez-Amado CJ, Langley E, Hidalgo-Miranda A (2018) Overview of mitochondrial germline variants and mutations in human disease: Focus on breast cancer (Review). International Journal of Oncology 53: 923-936.

Jones ME, Schoemaker MJ, Wright LB, Ashworth A, Swerdlow AJ (2017) Smoking and risk of breast cancer in the Generations Study cohort. Breast Cancer Research 19:118 DOI 10.1186/s13058-017-0908-4.

Keller G, Vogelsang H, Becker I et al., (1999) Diffuse type gastric and lobular breast carcinoma in a familial gastric cancer patient with an E-cadherin germline mutation. American Journal of Pathology 155(2):337–342.

Khalis M, Charbotel B, Chajès V, Rinaldi S, Moskal A, Biessy C, Dossus L, Huybrechts I et al., (2018) Menstrual and reproductive factors and risk of breast cancer: A case-control study in the Fez region, Morocco. PLoS ONE 13(1): e0191333.

Khodarahmi M, Azadbakht L (2014). The association between different kinds of fat intake and breast cancer risk in women. International journal of preventive medicine (5)1:6-15

Kim A, Lundgreen A, Wolff RK, Fejerman L, John EM, Torres-Mejía G, Ingles SA, Boone SD et al., (2016) Red meat, poultry, and fish intake and breast cancer risk among Hispanic and Non-Hispanic white women: The Breast Cancer Health Disparities Study. Cancer Causes and Control 27(4): 527–543

Kim JH, Lee J, Jung SY, Kim J (2017) Dietary Factors and Female Breast Cancer Risk: A Prospective Cohort Study. Nutrient 9:1331. doi.org/10.3390/nu9121331

Kloss-Brandsta□tter A, Pacher D, Scho□nherr S, Weissensteiner H, Binna R, Specht G, Kronenberg F (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Human Mutation 32(1):25-32.

Kohane, I. S., Masys, D. R., and Altman, R. B. (2006). The incidentalome: a threat to genomic medicine. Journal of the American Medical Association 296:212–215.

Kooshki A, Moghaddam MY, Akbarzadeh R (2016) Study of fruit and vegetable intake in breast cancer patients in the city of Sabzevar. Electronic Physician 9(8):3011-3014.

Kotsopoulos J, Lubinski J, Salmena L, Lynch HT, Kim-Sing C, Foulkes WD, Ghadirian P, Neuhausen SL et al., (2012) Breastfeeding and the risk of breast cancer in BRCA1 and BRCA2 mutation carriers. Breast Cancer Research 14:R42

Kumar R, Indrayan A (2011) Receiver Operating Characteristic (ROC) Curve for Medical Researchers. Indian Pediatrics 48:17.

Kumar V, Abbas AK, Fausto N, Aster JC (2010) Pathogenics Basics of Disease, Robbins and Cotrans, Elsevier, 8e, ISBN: 978-81-312-2491-5.

Kuzhan A, Adli M, Alkis HE, Caglayan D (2013) Hormone receptor and HER2 status in patients with breast cancer by races in southeastern Turkey. Journal of the Balkan Union of Oncology 18(3):619-622.

Larsen MJ, Thomassen M, Gerdes AM, Kruse TA (2014) Hereditary breast cancer: clinical, pathological and molecular characteristics. Breast Cancer 8:145-55.

Li H (2014) Toward better understanding of artifacts in variant calling from high-coverage samples. Bioinformatics. 30(20):2843–51.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R. 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9.

Li H. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. 2013. arXiv:1303.3997v1 [q-bio.GN].

Li J, Jing R, Wei H, Wang M, Xiaowei Q, Liu H, Jian L, Ou JH et al., (2018) Germline mutations in 40 cancer susceptibility genes among Chinese patients with high hereditary risk breast cancer. International Journal of Cancer 144:281–289.

Li L, Chen L, Li J, Zhang W, Liao Y, Chen J, Sun Z (2016) Correlational study on mitochondrial DNA mutations as potential risk factors in breast cancer. Oncotarget 7(21).

Lodha R, Joshi A, Paul D, Lodha KM, Nahar N, Shrivastava A, Bhagat VK, Nandeshwar S (2011) Association between reproductive factors and breast cancer in an urban set up at central India: A case-control study. Indian Journal of Cancer 48:3.

Loprinzi PD, Cardinal BJ, Smit E, Winters-Stone KM (2012) Physical activity and breast cancer risk. Journal of Exercise Science & Fitness 10:1e7.

Lu C, Sun H, Huang J, Yin S, Hou W, Zhang J, Wang Y, Xu Y, Xu H (2017) Long-Term Sleep Duration as a Risk Factor for Breast Cancer: Evidence from a Systematic Review and Dose-Response Meta-Analysis. Biomed Research International doi.org/10.1155/2017/4845059.

Mahdavi M, Nassiri M, Kooshyar MM, Vakili- Azghandi M, Avan A, Sandry R, Pillai S,Lam AKY, Gopalan V (2019) Hereditary breast cancer; Genetic penetrance and current status with BRCA. J Cell Physiol 234:5741–5750.

Malvia S, Bagadi SA, Dubey US, Saxena S (2017) Epidemiology of breast cancer in Indian women. Asia-Pacific Journal of Clinical Oncology 13: 289–295.

Mangtani P, Maringe C, Rachet B, Coleman MP, Silva IDS (2010) Cancer mortality in ethnic South Asian migrants in England and Wales (1993-2003): patterns in the overall population and in first and subsequent generations. British Journal of Cancer 102:1438-43.

Maskarinec G, Fontaine A, Torfadottir JE, Lipscombe LL, Lega IC, Figueroa J, Wild S (2018) The Relation of Type 2 Diabetes and Breast Cancer Incidence in Asian, Hispanic and African American Populations—A Review. Canadian Journal of Diabetes 42 (2018) 100–105.

Mavaddat N, Antoniou AC, Easton DF, Garcia-Closas M (2010) Genetic susceptibility to breast cancer. Molecular oncology 4(3):174–191.

Mavaddat N, Pharoah PD, Blows F, Driver KE, Provenzano E, Thompson D, MacInnis RJ, Shah M et al., (2010) Familial relative risks for breast cancer by pathological subtype: a population-based cohort study. Breast Cancer Research, 12:R10.

McDonald JA, Goyal A, Terry MB (2013) Alcohol Intake and Breast Cancer Risk: Weighing the Overall Evidence. Current Breast Cancer Reports 5(3) doi:10.1007/s12609-013-0114-z.

McMahon S, LaFramboise T (2014) Mutational patterns in the breast cancer mitochondrial genome, with clinical correlates. Carcinogenesis 35(5):1046–1054.

Melvin JC, Wulaningsih W, Hana Z, Purushotham AD, Pinder SE, Fentiman I, Gillett C, Mera A et al., (2016) Family history of breast cancer and its association with disease severity and mortality. Cancer Medicine 5(5):942–949.

Miao L, Yu Y, Ji Y, Zhang B, Yuan Z, Du Y, Zhu L, Wang R et al., (2017) Association between BRCA1 P871L polymorphism and cancer risk: evidence from a meta-analysis. Oncotarget 8(18):30587-30594

Miki Y, Swensen J, Shattuck-Eidens D, Futreal PA, Harshman K, Tavtigian S, Liu Q, Cichran C, Bennett LM, Ding W, et al. (1994) A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266: 66-71

Mirunalini S, Dhamodharan G, Karthishwaran K, Sugunadevi G (2010) Correlation of Estrogen and Progesterone Receptors Status with the Grade and Type of Breast Cancer. Research Journal of Agriculture & Biological Sciences 6(3):199-203.

Muhammad N, Nawaz R, Khan FA, Naeemi H, Rashid MU (2018) Inherited Genetic Susceptibility to Breast Cancer in Pakistan. Journal of Cancer and Allied Specialties 4(2):2.

Nadiah TB, Hasnan J, Zafarina Z (2012) Association of Mitochondrial DNA 10398 Polymorphism in Invasive Breast Cancer in Malay Population of Peninsular Malaysia. Malaysian Journal of Medical Sciences 19(1): 36-42.

Nichols KE, Malkin D, Garber JE, Fraumeni JF, Li FP (2001) Germ-line p53 mutations predispose to a wide spectrum of early-onset cancers. Cancer Epidemiology Biomarkers and Prevention 10(2):83–87.

Noordiana N, Fatimah AB, Farhana YCB (2011) Formaldehyde content and quality characteristics of selected fish and seafood from wet markets. International Food Research Journal 18: 125-136.

Opdahl S, Alsaker MD, Janszky I, Romundstad PR, Vatten LJ (2011) Joint effects of nulliparity and other breast cancer risk factors. British Journal of Cancer. 23;105(5):731-6.

Palachandra A, Ishawaraprasad GD, Sreelatha CY, Sumana M (2017) Risk factors associated with carcinoma breast: a case control study International Surgery 4(9):3136-3140.

Palodhi A, Ghosh S, Biswas NK, Basu A, Majumder PP, Maitra A et al., (2018) Profiling of genomic alterations of mitochondrial DNA in gingivobuccal oral squamous cell carcinoma: Implications for disease progress. Mitochondrion doi:10.1016/ j.mito.2018.09.006

Parada JrH, Steck SE, Bradshaw PT, Engel LS, Conway K, Teitelbaum SL, Neugut AI, Santella RM, Gammon MD (2017) Grilled, Barbecued, and Smoked Meat Intake and Survival Following Breast Cancer. Journal of National Cancer Institute 109(6)

Parr RL, Dakubo GD, Thayer RE, McKenney K, Birch-Machin MA (2006) Mitochondrial DNA as a potential tool for early cancer detection. Human Genomics. 2(4):252–257.

Parrella P, Xiao Y, Fliss M, Sanchez-Cespedes M, Mazzarelli P, Rinaldi M, Nicol T, Gabrielson E et al., (2001) Detection of Mitochondrial DNA Mutations in Primary Breast Cancer and Fine-Needle Aspirates. Cancer Research 61:7623–7626.

Pharoah PDP, Guilford P, Caldas C (2001) Incidence of gastric cancer and breast cancer in CDH1 (E-cadherin) mutation carriers from hereditary diffuse gastric cancer families. Gastroenterology 121(6):1348–1353.

Pourzand A, Fakhree MBA, Hashemzadeh S, Halimi M, Daryani A (2011) A Hormone Receptor Status in Breast Cancer and its Relation to Age and Other Prognostic Factors Breast Cancer. Basic and Clinical Research 5 87–92.

Rahman N, Seal S, Thompson D, Kelly P, Renwick A, Elliott A, Reid S, Spanova K et al., (2007) PALB2, which encodes a BRCA2-interacting protein, is a breast cancer susceptibility gene. Nature Genetics 39:165–167.

Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Ichihara S et al., (2010) Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Research 12: 207 doi:10.1186/bcr2607.

Richards S, Aziz N, Bale S, Bick D, Soma Das, Gastier-Foster J, Grody WW, Hegde M et al., (2015) Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. Genetics in Medicine 17:5.

Ricks-Santi L, McDonald JT, Gold B, Dean M, Thompson N, Abbas M, Wilson B, Kanaan Y et al., (2017) Next generation sequencing reveals high prevalence of Brca1 and Brca2 variants of unknown significance in early-onset breast cancer in African American Women. Ethnicity & Disease 27:2

Rossing M, Sørensen CS, Ejlertsen B, Nielsen FC (2019) Whole Genome Sequencing of Breast Cancer. Journal of Pathology, Microbiology and Immunology, Acta Pathologica, Microbiologica, et Immunologica Scandinavica, 127: 303–315.

Sandhu GS, Erqou S, Patterson H, Mathew A (2016) Prevalence of Triple-Negative Breast Cancer in India: Systematic Review and Meta-Analysis. Journal of Global Oncology 2:6.

Saxena S, Kaushal M, Sharma J, Zomawia E, Kapur S (2010) Genomic alteraions in breast cancer patients from Northeast India using 10K SNP arrays. Genome Biology 11(34).

Schumacher V, Vogel T, Leube B, Driemel C, Goecke T, Möslein G, Royer-Pokora B (2005) STK11 genotyping and cancer risk in Peutz-Jeghers syndrome. Journal of Medical Genetics 42:428–435.

Schwartz AM, Henson DE, Chen D, Rajamarthandan S (2014) Histologic Grade Remains a Prognostic Factor for Breast Cancer Regardless of the Number of Positive Lymph Nodes and Tumor Size, A Study of 161 708 Cases of Breast Cancer From the SEER Program. Archives of Pathology and Laboratory Medicine -138.

Shahi RB, Brakeleer SD, Caljon B, Pauwels I, Bonduelle M, Joris S, Fontaine C, Vanhoeij M, Dooren SV, Teugels E, De Grève J (2019) Identification of candidate cancer predisposing variants by performing whole-exome sequencing on index patients from BRCA1 and BRCA2-negative breast cancer families. BCM Cancer 9:313.

Sharma-Oates A, Shaaban AM, Tomlinson I, Wynne L, Cazier JB, Sundar S (2018) Heterogeneity of germline variants in high risk breast and ovarian cancer susceptibility genes in India. Precision Clinical Medicine 1(2):75–87.

Sheikh A, Hussain SA, Ghori Q, Naeem N, Fazil A, Giri S, Sathian B, Mainali P et al., (2015) The Spectrum of Genetic Mutations in Breast Cancer. Asian Pacific Journal of Cancer Prevention 16 (6), 2177-2185.

Shiovitz S, Korde LA (2015) Genetics of breast cancer: a topic in evolution. Annals of Oncology 26:1291–1299.

Siadati S, Sharbatdaran M, Nikbakhsh N, Ghaemian N (2015) Correlation of ER, PR and HER-2/Neu with other Prognostic Factors in Infiltrating Ductal Carcinoma of Breast. Iran Journal of Pathology 10(3):221-226.

Singh J, Thota N, Singh S, Padhi S, Mohan P, Deshwal S, Sur S, Ghosh M et al., (2018) Screening of over 1000 Indian patients with breast and/or ovarian cancer with a multi-gene panel: prevalence of BRCA1/2 and non-BRCA mutations; Breast Cancer Research and Treatment 170:189–196.

Singh JK, Nath A, Mishra M, Jain P, Kumar A, Ali M, Kumar R (2011) Early marriage and premenopause as a risk factor for breast cancer. Human Physiology 38 (2011) 4232-4234.

Sirisena ND, Adeyemo A, Kuruppu AI, Neththikumara N, Samaranayake N, Dissanayake VHW (2018) Genetic determinants of sporadic breast cancer in Sri Lankan women. BMC Cancer 18:180.

Sirisena ND, Adeyemo A, Kuruppu AL, Neththikumara N, Samaranayake N, and Dissanayake VHW (2018) Genetic determinants of sporadic breast cancer in Sri Lankan women. BMC Cancer 18:180 doi.org/10.1186/s12885-018-4112-4

Stookey JD, Belderson PE, Russell JM, Barker ME (1997) Relationship of food groups and water intake to colon cancer risk. Cancer Epidemiology, Biomarkers & Prevention 6:657-658.

Sultana R, Ahmmed R, Khan MJ, Hoque MS, Sabuj AA (2018) Quality Analysis of Fruits, Vegetables and Fish Available in Local Market of Southern Districts in Bangladesh. Duet Journal 4(1).

Szabo CI, Wagner LA, Francisco LV, Roach JC, Argonza1 R, King M, Elaine A, Ostrander E.A (1996) Human, canine and murine BRCA1 genes: sequence comparison among species. Human Molecular Genetics 5(9):1289–1298.

Tazzite A, Jouhadi H, Saiss K, Benider A, Nadifi S (2013) Relationship between family history of breast cancer and clinicopathological features in Moroccan patients. Ethiopian Journal of Health Sciences 23:2.

Thiygarajan M, Navrathan N, Mohanapriya T, Kumar A, Singh B (2015) Correlation between estrogen receptor, progesterone receptor, HER- 2/neu status and other prognostic factors in carcinoma breast in Indian population. International Surgery Journal 2(4):515-522.

Thomassen M, Hansen TVO, Borg A, Lianee HT, Wikman F, Pedersen IS, Bisgaard ML, Nielsen FC et al., (2008) BRCA1 and BRCA2 mutations in Danish families with hereditary breast and/or ovarian cancer. Acta Oncologica 47(4), 772-777.

Thyagarajan B, Wang R, Nelson H, Barcelo H, Koh W-P, Yuan J-M (2013) Mitochondrial DNA Copy Number Is Associated with Breast Cancer Risk. PLoS ONE 8(6): e65968.

Tommasi S, Favia P, Weigl S, Bianco A, Pilato B, Russo L, Paradiso A, Petruzzella V (2014) Mitochondrial DNA variants and risk of familial breast cancer: An exploratory study. International Journal of Oncology 44:1691-1698

Uddin R, Wahid MI, Jesmeen T, Huda NH, Sutradhar KB (2011) Detection of Formalin in Fish Samples Collected from Dhaka City, Bangladesh. Stamford Journal of Pharmaceutical Sciences 4(1): 49-52.

Vallon-Christersson J, Cayanan C, Haraldsson C, Loman N, Bergthorsson JT, Brøndum-Nielsen K, Gerdes A, Møller P et al., (2001) Functional Analysis of BRCA1 C-Terminal Missense Mutations Identified in Breast and Ovarian Cancer Families. Human Molecular Genetics 10(4):353–360.

Venselaar H1, Te Beek TA, Kuipers RK, Hekkelman ML, Vriend G. (2010) Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. BMC Bioinformatics. 2010 Nov 8;11(1):548.

Walsh T, Lee MK, Casadei S, Thornton AM, Stray SM, Pennil C, Nord AS, Mandell JB et al., (2010) Detection of inherited mutations for breast and ovarian cancer using genomic capture and massively parallel sequencing. Proceedings of the National Academy of Sciences 107(28):12631.

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

Wang XX, Jiang YZ, Li JJ, Song CG, Shao ZM (2016) Effect of nodal status on clinical outcomes of triple-negative breast cancer: a population-based study using the SEER 18 database. Oncotarget 7:29.

Weigl S, Paradiso A, Tommasi S (2013) Mitochondria and Familial Predisposition to Breast Cancer. Current Genomics 14(3):195-203.

Wong NS, Anderson BO, Khoo KS, Ang PT, Yip CH, Lu YS, Voravud N, Shao ZM, Pritchard KI (2009) Management of HER2-positive breast cancer in Asia: consensus statement from Asian Oncology Summit. Lancet Oncology 10:1077-85.

Yin Y, Shen WH (2008) PTEN: a new guardian of the genome. Oncogene 27:5443–5453.

Zhang X, Morera S, Bates PA, Whitehead PC, Coffer AI, Hainbucher K, Nash RA, Sternberg MJ et al., (1998) Structure of an XRCC1 BRCT domain: a new protein-protein interaction module. European Molecular Biology Organization Journal 17:6404–6411.

Zhang Y, Cai Q, Xiao-Ou Shu XO, Gao YT, Li C, Zheng W, Jirong Long J (2015) Whole-Exome Sequencing Identifies Novel Somatic Mutations in Chinese Breast Cancer Patients. Journal of Molecular Genetics Med 9:4.

Zheng K, Tan JX, Li F, Li HY, Zeng XH et al. (2018) Clinicopathologic Factors Related to the Histological Tumor Grade of Breast Cancer in Western China: An Epidemiological Multicenter Study of 8619 Female Patients. Translational oncology 11(4):1023-1033.

Zhu W, Qin W, Bradley P, Wessel A, Puckett CL, Sauter ER (2005) Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid Carcinogenesis 26(1):145-152.

Zhu WQW, Bradley P, Wessel A, Puckett CL, Sauter ER (2005) Mitochondrial DNA mutations in breast cancer tissue and in matched nipple aspirate fluid. Carcinogenesis 26(1):145–152

**Website/ Internet Sources**

Cancer India, http://cancerindia.org.in/globocan-2018-india-factsheet/

Centers for Disease Control Prevention, What Is Breast Cancer? https://www.cdc.gov/cancer/breast/basic_info/what-is-breast-cancer.htm

ClinVar. https://www.ncbi.nlm.nih.gov/clinvar/variation/427022/ (Accessed Aug 21, 2019)

Environmental Defense Fund, The benefits of eating fish http://seafood.edf.org/benefits-eating-fish (Accessed 11. Dec. 2019)

Globocan. International Agency for Research on Cancer, 2018 http://gco.iarc.fr/today/data/factsheets/cancers/20-Breast-fact-sheet.pdf

Homer N, Merriman B. TMAP: the Torrent Mapping Alignment Program. https://github .com/io ntorrent/TMAP

http://www.ncdirindia.org/NCRP/ALL_NCRP_REPORTS/PBCR_REPORT_2012_2014/ALL_CONTENT/PDF_Printed_Version/Chapter2_Printed.pdf

Human Anatomy, Picture of the Breasts, https://www.webmd.com/women/picture-of-the-breasts#1

National Cancer Institute, NIH- What Is Cancer?, https://www.cancer.gov/about-cancer/understanding/what-is-cancer

National Cancer Institute, NIH- What Is Cancer?, https://www.cancer.gov/about-cancer/understanding/what-is-cancer

The Wire, Mizoram Has Highest Number of Inter-Caste Marriages in India With 55%. https://thewire.in/uncategorised/mizoram-has-highest-number-of-inter-caste-marriages-in-india-with-55

Wake Forest University School Of Medicine. "Use Of Smokeless Tobacco May Lead To Breast Cancer, Wake Forest Team Reports." ScienceDaily, 2000. www.sciencedaily.com/releases/2000/05/000509003313.htm

# *Appendices*

## Appendix I: Papers published in peer reviewed journals

- **Doris Zodinpuii**, Jeremy Lalrinsanga Pautu, Bawitlung Zothankima, Lalawmpuii Pachuau, Nachimuthu Senthil Kumar (2019) 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***. doi.org/10.1007/s11356-019-06944-8.
- Madathil S, Senthil Kumar N, **Zodinpuii D**, Muthukumaran RB, Lalmuanpuii R, Nicolau B (2018) Tuibur: tobacco in a bottle-commercial production of tobacco smoke-saturated aqueous concentrate. ***Addiction*** 113(3):577-580. Doi:10.1111/add.14117.

## Manuscript under review

- **Doris Zodinpuii**, Jeremy Lalrinsanga Pautu, Bawitlung Zothankima, Doris Lallawmzuali, Ashok Kumar Varma, Nachimuthu Senthil Kumar. A landscape of Mitochondrial DNA mutations and their risk towards Breast Tumorigenesis. ***Mitochondrion.***
- **Doris Zodinpuii**, Jeremy L. Pautu, B. Zothankima, Doris Lallawmzuali, Ashok K. Varma, Nachimuthu Senthil Kumar. Breast Cancer is significantly associated with Life Style Diseases in a Northeast Indian Tribal Population. ***Ethnicity and Health.***
- **Doris Zodinpuii**, Jeremy L. Pautu, B. Zothankima, Doris Lallawmzuali, Ashok K. Varma, Nachimuthu Senthil Kumar. Breast cancer susceptibility gene mutations in Mizo population. ***Biomarkers.***
- **Doris Zodinpuii**, Jeremy L. Pautu, B. Zothankima, Doris Lallawmzuali, Pradnya Kotwal, Rajiv Sarin, Ashok K. Varma, Nachimuthu Senthil Kumar. Two Rare Variants, C376T:p.Q126X and T5089C:p.C1697R are Found in Breast Cancer Susceptibility Gene BRCA1. ***Current Genomics***

## Appendix II: List of presentation in conference/symposium/seminar/trainings

- Presented a paper on "Impact of Epidemiological Factors Associated with Breast Cancer in Mizo Women " in National Conference on Recent Advances in Biotechnology organized by Department of Biotechnology, School of Life Sciences, Mizoram University during 9 & 10th Nov, 2017.
- Participated as a Resource person in the National Workshop on "Hands on Training on DNA Barcoding and Phylogenetics", organized by Advanced level State Biotech- Hub Facility, Department of Biotechnology, Mizoram University during 20-25th March 2017.

## Appendix III: List of seminar/symposium/conference/workshops attended

- Participated in workshop on "Bioinformatics- Structure & Determination of macromolecules" organized by Department of Biotechnology, Mizoram University held on 28 & 29th March, 2011 at Bioinformatics Infrastructure Facility, Department of Biotechnology, Mizoram University.
- Participated in "Workshop on Capacity Building in Effective Management of Intellectual Property Rights (IPRs) in Biotechnology by Universities and Research Institutes in Mizoram " organized by Biotech Consortium India Limited (BCIL), New Delhi, held at Mizoram University during 27 & 28th Aug, 2014.
- Participated in workshop organized by Bioinformatics Infrastructure Facility, Department of Biotechnology, Mizoram University and Schrodinger, Bangalore on "Molecular Docking and Virtual Screening" held on 2-4 Oct, 2014 at Department of Biotechnology, Mizoram University.
- Participated in workshop on "Statistical and Computing Methods for Life-Science Data Analysis" organized by the Biological Anthropology Unit, Indian Statistical Institute, Kolkata during 9-16th Feb, 2015 at Department of Environmental Science, Pachhunga University College, Aizawl, Mizoram.
- Participated in North-East Autumn School on Human Genetics: Techniques and Statistical Analyses organized by Indian Statistical Institute, Kolkata & Mizoram University during 8-11th Sept, 2015.
- Participated in Research Training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics- Driven Approaches" held at Department of Molecular Biology & Bioechnology, Tezpur University during 9-13th May, 2016.
- Participated in Advanced Research Training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics- Driven Approaches" held at National Institute of Biomedical Genomics Kalyani from 27 Feb- 10 March, 2017.
- Participated in "Science Communication Workshop (SciComm 101)" organized by The Wellcome Trust/DBT India Alliance at Mizoram University on 6th June 2017.

- Participated in workshop organized by Civil Hospitals, Aizawl and Mizoram University on "The Concept and Application of Genomics in Clinical Medicine" held on 11 Aug, 2018 at Civil Hospital, Aizawl, Mizoram.
- Participated in the 12th Annual Convention of Association of Biotechnology and Pharmacy (ABAP) & International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Trends (BEHIET), organized at the School of Life Science, Mizoram University, Aizawl, Mizoram during 12 to 14 Nov, 2018.
- Participated in National Workshop organized by Bioinformatics Infrastructure Facility (BIF) on "A brief introduction to Bioinformatics and Systems Biology" held during 13th and 14th Dec, 2018 at Department of Biotechnology, Mizoram University.

ADDICTION

SSA SOCIETY FOR THE STUDY OF ADDICTION

Edited By: Editor-in-Chief: Robert West

Impact factor: 6.851

ISI Journal Citation Reports © Ranking: 2018: 12/146 (Psychiatry); 8/142 (Psychiatry (Social Science)); 1/19 (Substance Abuse); 2/35 (Substance Abuse (Social Science))

Online ISSN: 1360-0443




LATEST ISSUE >

Volume 114, Issue 12
December 2019



# About this journal

*Addiction* is the official journal of the Society for the Study of Addiction, and has been in publication since 1884.
The journal publishes peer-reviewed research reports on pharmalogical and behavioural addictions, bringing together research conducted within many different disciplines.

**Welcome to the new look Wiley Online Library.**

## Featured


### Virtual Issues

Read the latest themed virtual issues and online collections from Addiction:

Submit an Article

Browse free sample issue

Get content alerts

Subscribe to this journal

Published on behalf of the Society for the Study of Addiction

SSA SOCIETY FOR THE STUDY OF ADDICTION

# BIODATA

## ***Personal Information***

| | | |
|---|---|---|
| Name | : | Doris Zodinpuii |
| Date of Birth | : | 31st May 1988 |
| Permanent Address | : | Hno. C-15/A, Ramhlun North, Aizawl, Mizoram |
| Contact No. | : | 8837466743 |
| Email id | : | curio7doris@gmail.com |

## ***Qualification***

| EXAM PASSED | INSTITUTION | BOARD/ UNIVERSITY | PERCENTAGE (%) |
|---|---|---|---|
| HSLC | Mamawii HSS | M.B.S.E | 60.6 |
| AISSCE | M.I.C.E | C.B.S.E | 57 |
| Bsc. Biotechnology | S.K Women College | M.U | 67.86 |
| Msc. Biotechnology | M.Z.U | M.Z.U | 68.56 |

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-019-06944-8

RESEARCH ARTICLE

# 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

Doris Zodinpuii[1] · Jeremy Lalrinsanga Pautu[2] · Bawitlung Zothankima[3] · Lalawmpuii Pachuau[4] · Nachimuthu Senthil Kumar[1]

Received: 3 May 2019 / Accepted: 1 November 2019


## Abstract

The aim was to assess the association of histological tumor grade with other clinical features and epidemiological factors of women with invasive breast carcinoma. A retrospective study of 103 Mizo breast cancer patients visiting hospitals was made in Aizawl, Mizoram, Northeast India. With a prior consent, information on epidemiological factors and family history in relation to cancer was obtained. Clinical reports were obtained from their medical records. The frequency of distribution was calculated for age at diagnosis and tumor characteristics. Statistical analysis for different variables was done using a chi-square test. $p < 0.05$ was considered significant. The histological tumor grades in our studies were found to be associated with lymph node invasion ($p < 0.021$), different subtype of hormone receptor such as ER status ($p < 0.004$), ER/PR status ($p < 0.007$), HER2/neu status ($p < 0.014$), and ER/PR/HER2 status ($p < 0.025$). A patient with a family history of breast cancer in their 1st degree relative is also seen to have association in determining the tumor grade ($p < 0.003$). Reproductive history, lifestyle and dietary habits, tobacco, and alcohol consumption were found to have no influence on breast cancer tumor grade. Our results showing significant correlation between status of lymph node, ER, PR, and HER2/neu oncoprotein and family history with 1st degree relative breast cancer are the first time report to target and focus on the possible role of biomarkers for diagnosis among the Mizo tribal breast cancer patients.

**Keywords** Breast cancer · Hormone receptor · HER2/neu oncoprotein · Tumor grade · Lymph node · Tobacco · Epidemiology · Family history

## Introduction

Mizoram, situated in the northeastern part of India with not more than 13 lakhs population with a total area of 21,081 $km^2$, is known for three striking records. The first good record made was in socio-economic progress for having the highest literacy rate in India with 90.27% (Hindustan times 2006). Second is for being the state with highest prevalence for tobacco usage (67.2%) in India with males and females (72.5% and 61.6%, respectively) consuming tobacco in different forms. Exposure to secondhand smoke inside the house is very high (97.7%). Lastly, Mizoram has the highest incidence of cancers, and breast is one of the leading sites among the women (Mizoram Population Census 2018). However, being breast as one of the most prevalent cancers, the variants and their frequency and risk factors to breast cancer (BC) in this Mongoloid population also remain unexplored and may be applicable to other populations as well.

Responsible editor: Philippe Garrigues

**Electronic supplementary material** The online version of this article (https://doi.org/10.1007/s11356-019-06944-8) contains supplementary material, which is available to authorized users.

✉ Nachimuthu Senthil Kumar
nskmzu@gmail.com

[1] Department of Biotechnology, Mizoram University, Tanhril, Aizawl, Mizoram 796009, India

[2] Department of Oncology, Mizoram State Cancer Institute, Zemabawk, Aizawl, Mizoram, India

[3] Department of Radiation Oncology, Mizoram State Cancer Institute, Zemabawk, Aizawl, Mizoram, India

[4] Department of Pathology, Civil Hospital, Aizawl, Mizoram, India

Published online: 25 November 2019

BC is a key threat and major cause of cancer-related death among women worldwide. The age range of women suffering from BC in Mizoram is as wide as 19–91 years old, and the incidence of BC ranks the third highest with 13.0%, following cervix uteri (15.9%) and lung cancer (15.6%) (NCRP 2014). It has been expected to witness an increase in the occurrence of BC cases within these few years to a certain fold. Like all other types of cancer, BC is a complex heterogeneous disease regulated by a multitude of genetic and epigenetic alterations (Atif et al. 2018). Regional and ethnic variations in the incidence and continuous increase in the figure of BC cases cannot be explained by gene functions caused by inherited or acquired mutations. Many studies have reported a significant correlation between BC increase risk and several epidemiological factors (Perez-Solis et al. 2016).

Several clinical and pathological factors, which include tumor size, histological tumor grade, hormone receptor (Estrogen ER and Progesterone PR) status, and lymph node metastasis, affect the prognosis of BC (Kuzhan et al. 2013). One of the momentous developments in the assessment of BC has been the conception that the presence of hormone receptors in the tumor tissue correlates well with response to chemotherapy and hormone therapy (Mirunalini et al. 2010). Lymph nodes metastasis indicates advanced disease stage suggesting that cancer cells have extent to distant location. At the time of diagnosis, ~ 50% of all BCs already spread to the sentinel lymph node (Pourzand et al. 2011).

Even though, in BC, the most well-established prognostic factor is histological tumor grade (Rakha et al. 2010); not much report is available to explore whether any correlation exists between the tumor grade with other clinical features, family history, as well as dietary, and life style habits. Association between tumor grade with tobacco-related products that are very much common among the local people such as, Betel nut/pan, gutkha (mixture of tobacco and betel nut), sahdah and khaini (processed tobacco leaves that are stuffed in between the gums), tuibur (liquid form of tobacco), and cigarette also remains unknown (Madathil et al. 2018). The objective of this study is to find out the association of tumor grade with epidemiological factors viz. reproductive history, lifestyle and dietary habits, tobacco and alcohol consumption, family history of breast and other cancers, and immunehistochemical markers such as hormone receptor and HER2/neu oncoprotein among the Mizo tribal BC patients' community in Northeast India.

## Methods

### Patient and data selection

This is a retrospective study of only Mizo ethnic BC patients of any age within the state of Mizoram, registered in Mizoram State Cancer Institute (MSCI) and Civil Hospital, Aizawl, Mizoram, Northeast India between the years 2013 and 2017. Patients not registered in both the hospitals and undergoing treatment outside Mizoram were also included in the study. Information were obtained in a two-way approach: the first is by going to hospitals and interviewing the patients and the second by collecting the name, contact number, and home address of all BC patients from different hospitals in Aizawl City, which also include the list from hospital based cancer registry (HBCR), and the patients were interviewed telephonically. And in case patients could not be reached by telephones, house visit was made and interview was done in person. In both approaches, the patient's consent was first taken and their details on epidemiological factors and family history in relation to cancers were obtained. The histopathology reports were retrieved from their medical files and also from the hospitals they visited, and data on histological type, grade of tumors, lymph node involvement, hormone receptors, and HER2 receptor status was collected using a standard pro forma. The histopathological grading of the BC tumor was done at the Department of Pathology in Mizoram State Cancer Institute and Civil Hospital, Aizawl, Mizoram according to the Nottingham modification of the Bloom Richardson grading system accepted by the World Health Organization (Rakha et al. 2010; Kaur et al. 2016). Questionnaires and histological reports of all the samples were screened for the completeness of information, and the inconsistent data were removed. A total of 426 patients were interviewed, and only 103 cases met the complete required information and were included in the analysis. In order to observe the best possible association between tumor grades with the factors studied, we further analyzed with only 51 cases that included the highest incidence age range of 40–55 years.

### Classification of factor variables

Clinical features such as lymph node status as well as tumor characteristics such as tumor type (infiltrating ductal carcinoma and infiltrating lobular carcinoma), tumor side (Left, right or both), tumor site (all quadrant, retro-areolar, lower inner/outer quadrant, upper inner/outer quadrant), and tumor sizes were recorded. Lymph node was divided into positive invasion and negative invasion. Type, side, and site of tumor were grouped into their following category, while tumor size was divided into three size groups as 1–5 cm, 5–10 cm, and > 10 cm. Hormone receptors (ER and PR) and HER2/neu onco protein were analyzed individually as well as in different subtypes.

Factors in lifestyle (exercise, sleeping hours, and night duty) and dietary habits (pork, fish, chicken, beef, fruits, vegetables, Sa-um (fermented pork fats), smoked meat, smoked vegetables, salt, oil, and water) were considered on a weekly basis. Other factors such as salt and oil consumption, sleeping

hours, night duty, and water intake were recorded on a daily basis. Tobacco and tobacco-related products (Betel nut/pan, gutkha, sahdah, khaini, tuibur, and cigarette) and alcohol were also divided into the amount of consumption from the start of their addiction on a daily basis.

Reproductive histories (age at menarche, age at marriage, age at first delivery, number of live birth, lactation duration, age at diagnosis, marital status, menopausal status, parity, breast feeding, use of oral contraceptive pills, and abortion) were categorized in different age range and duration depending on the factors considered. In case of factors such as age at marriage, number of live birth, and lactation duration, samples that have not gone through marriage, child delivery, or lactation were not included in the analysis. Family history in relation to breast and other cancers was categorized into no relatives and with the number of relatives they have for each cancer.

### Statistical analysis

Statistical analyses were conducted using the IBM Statistical Package for Social Sciences (SPSS), software version 22.0 for Windows. The frequency distribution was calculated for age at diagnosis and tumor characteristics. Statistical analysis for different variables was done using a chi-square test. Since the analysis involves multiple subgroups, significance was set at 5% level ($p < 0.05$) (Kuzhan et al. 2013; Pourzand et al. 2011).

## Results

In our study, the patients' age range is 23–67 years with mean age at diagnosis of 48 years. The distributions of clinical and pathological features of these103 cases are summarized in Supplementary Table 1. Infiltrating ductal carcinoma type of tumor is more common than that of infiltrating lobular carcinoma. Tumor size ranges from 1 to 20.5 cm, but tumor size less than 5 cm was most commonly observed (86.40%). Invasion of lymph node was observed in 52.42%, and the majority of the cases belong to grade I and II. Considering the ER+, PR+, and HER2/neu oncoprotein alone, almost half of the cases show expression with 54.36%, 49.51%, and 49.51%, respectively. Rates of ER+/PR+ is 46.60% and is most common in grade I tumor and ER−/PR− is 42.71% and is most common in grade II tumor; the remaining 10% of the cases are either ER+/PR− (7.76%) or ER−/PR+ (2.91%) and are observed only in grade I and II and are not found in grade III tumor. The most frequent triple subtype is ER+/PR+/HER2 − (27.18%) followed by ER−/PR−/HER2+ (25.24%). A total of 20 cases have triple positive ER+/PR+/HER2+ (19.41%), and 18 BC cases have triple negative ER−/PR−/HER2− (17.47%).

### Correlation between tumor grade with lymph node invasion, tumor characteristics, ER, PR, and HER2/neu oncoprotein

Lesser positive lymph node invasion was seen in grade I tumor, equal distributions in grade II tumor, but positive lymph node invasion is observed more in grade III tumor. The higher the grade of tumor, the more the cases of positive lymph node invasion were observed. Our analysis results also show that there is a significant association between the involvements of lymph node with grade of tumor ($p < 0.021$). No associations were seen between tumor grade with the type, side, size, and site of tumor. Association of tumor grade with hormone receptor status (estrogen and progesterone) and HER2/neu oncoprotein expression showed that the expression of ER ($p < 0.004$), Her2/neu ($p < 0.014$) alone, and expression of double subtype ER/PR ($p < 0.007$) and triple subtype ER/PR/HER2 ($p < 0.025$) have a significant association with grade of tumor. It has also been observed that the higher the grade of tumor the lesser the expression of Estrogen receptor (Table 1). In case of analysis performed for the same above factors by limiting age range to 40–55 years at the time of diagnosis, association of tumor grades was found to be significant with only tumor size while the rest of the other factors do not show any association (Supplementary Table 2).

### Correlation between tumor grade with reproductive and family history

We analyzed different variables of reproductive factors such as age at marriage, age at menarche, age at first delivery, number of live birth, duration of lactation, age at diagnosis, marital status, menopausal status, parity, breast feeding, use of oral contraceptive pills, and abortion to see whether they play a role in determining the tumor grade. None of the variables was found to have association with tumor grade. Age at diagnosis appears to be late; the majority of the cases (67.96%) developed BC at age greater than 40 years (Table 2). Analysis of the same reproductive factors, but limiting age range to 40–55 years at the time of diagnosis, also shows no significant association between tumor grades with reproductive history (Supplementary Table 3).

Self-reported family history of BC, ovarian cancer, any other type of cancer in their first and second degree relatives, and also acquired inheritable diseases was also analyzed to see their impact on tumor grade. Our analysis shows that with a $p < 0.003$, having at least one or more first degree relative with BC has a significant association in determining the grade of tumor. Interestingly, other variables like having a second degree relative with BC, first or second degree relatives with ovarian or other cancer, and whether the patients have inheritable disease do not show association with tumor grade (Table 2). Analysis was conducted for the

**Table 1** Correlation between tumor grade with lymph node invasion, tumor characteristics, ER, PR, and HER2/neu oncoprotein

| | Grade I (43) | Grade II (48) | Grade III (12) | *p* value |
|---|---|---|---|---|
| Lymph node invasion | | | | |
| (+ve) | 17 (39.53%) | 27 (56.25%) | 10 (83.33%) | 0.021 |
| (−ve) | 26 (60.46%) | 21 (43.75%) | 2 (16.67%) | |
| Tumor type | | | | |
| IDC | 42 (97.67%) | 48 (100.00%) | 11 (91.67%) | 0.169 |
| ILC | 1 (2.32%) | 0 (0.00%) | 1 (8.33%) | |
| Tumor type | | | | |
| Left | 17 (39.53%) | 30 (62.50%) | 5 (41.67%) | 0.228 |
| Right | 25 (58.13%) | 17 (35.42%) | 7 (58.33%) | |
| Both | 1 (2.32%) | 1 (2.08%) | 0 (0.00%) | |
| Tumor site | | | | |
| All quadrant | 0 (0.00%) | 2 (4.17%) | 1 (8.33%) | 0.700 |
| Retro-areola | 4 (9.30%) | 5 (10.42%) | 1 (8.33%) | |
| LIQ/LOQ/LIQ&LOQ | 9 (20.93%) | 10 (20.83%) | 1 (8.33%) | |
| UIQ/UOQ/UIQ&UOQ | 30 (69.76%) | 31 (64.58%) | 9 (75.00%) | |
| Tumor size (mm) | | | | |
| 1–5 | 37 (86.04%) | 42 (87.50%) | 10 (83.33%) | 0.440 |
| 5–10 | 6 (13.95%) | 5 (10.42%) | 1 (8.33%) | |
| > 10 | 0 (0.00%) | 1 (2.08%) | 1 (8.33%) | |
| ER+/− | | | | |
| +ve | 31 (72.09%) | 22 (45.83%) | 3 (25.00%) | 0.004 |
| −ve | 12 (27.91%) | 26 (54.17%) | 9 (75.00%) | |
| PR+/− | | | | |
| +ve | 24 (55.81%) | 24 (50.00%) | 3 (25.00%) | 0.168 |
| −ve | 19 (44.19%) | 24 (50.00%) | 9 (75.00%) | |
| HER2+/− | | | | |
| +ve | 14 (32.56%) | 30 (62.50%) | 7 (58.33%) | 0.014 |
| −ve | 29 (67.44%) | 18 (37.50%) | 5 (41.67%) | |
| Er/Pr status | | | | |
| ER+/PR+ | 24 (55.81%) | 21 (43.75%) | 3 (25.00%) | 0.007 |
| ER+/PR− | 7 (16.28%) | 1 (2.08%) | 0 (0.00%) | |
| ER−/PR+ | 0 (0.00%) | 3 (6.25%) | 0 (0.00%) | |
| ER−/PR− | 12 (27.91%) | 23 (47.92%) | 9 (75.00%) | |
| Er/Pr/Her2 status | | | | |
| Er+/Pr+/Her2+ | 6 (13.95%) | 12 (25.00%) | 2 (16.67%) | 0.025 |
| ER+/PR+/HER2− | 18 (41.86%) | 9 (18.75%) | 1 (8.33%) | |
| ER+/PR−/HER2+ | 3 (6.98%) | 0 (0.00%) | 0 (0.00%) | |
| ER+/PR−/HER2− | 4 (9.30%) | 1 (2.08%) | 0 (0.00%) | |
| ER−/PR+/HER2+ | 0 (0.00%) | 2 (4.17%) | 0 (0.00%) | |
| ER−/PR+/HER2− | 0 (0.00%) | 1 (2.08%) | 0 (0.00%) | |
| ER−/PR−/HER2+ | 5 (11.63%) | 16 (33.33%) | 5 (41.67%) | |
| ER−/PR−/HER2− | 7 (16.28%) | 7 (14.58%) | 4 (33.33%) | |

same family history and relation details by limiting age range to 40–55 years at the time of BC diagnosis. We found a consistent significant association between tumor grades with having at least one or more first degree relative with BC ($p < 0.029$). We also observed that, regardless of the samples considered for the analysis, no association was found between tumor grades with other factors included in the analysis (Supplementary Table 3).

## Correlation between tumor grade lifestyle and dietary habits, tobacco, and alcohol habits

We next determined the impact of lifestyle habits to see their association with tumor grade. Factors include the regularity of physical exercise, the duration of sleeping hours at night, and whether the patients' work involve night duty or not. According to our findings, exercise, sleeping hours, or night

**Table 2** Correlation between tumor grade with reproductive and family history of breast cancer

| | Grade I (43) | Grade II (48) | Grade III (12) | *p* value |
|---|---|---|---|---|
| Age range (mean) | 49 | 46 | 49 | 0.482 |
| Age at marriage (mean) | 24 | 23 | 22 | 0.164 |
| Age at menarche | 15 | 15 | 15 | 0.286 |
| Age at first delivery | 24 | 25 | 23 | 0.225 |
| No. of live birth | 3 | 4 | 3 | 0.321 |
| Duration of lactation (months) | 66 | 74 | 71 | 0.081 |
| Age at Diagnosis | | | | |
| < 40 | 12 (36.3%) | 16 (48.48%) | 5 (15.15%) | 0.642 |
| > 40 | 31 (44.28%) | 32 (45.71%) | 7 (10.00%) | |
| Marital status | | | | |
| Married | 39 (42.85%) | 40 (43.95%) | 12 (13.18%) | 0.225 |
| Single | 4 (33.33%) | 8 (66.66%) | 0 (0.00%) | |
| Menopausal | | | | |
| Pre | 13 (37.14%) | 18 (51.42%) | 4 (11.42%) | 0.765 |
| Post | 30 (44.11%) | 30 (44.11%) | 8 (11.76%) | |
| Parity | | | | |
| Parous | 38 (42.22%) | 40 (44.44%) | 12 (13.33%) | 0.211 |
| Nulliparous | 5 (38.46%) | 8 (61.53%) | 0 (0.00) | |
| Breast feeding | | | | |
| Yes | 38 (42.22%) | 40 (44.44%) | 12 (13.33%) | 0.289 |
| No | 5 (38.46%) | 8 (61.53%) | 0 (0.00%) | |
| Oral contraceptive pills | | | | |
| Yes | 9 (34.61%) | 15 (57.69%) | 2 (7.69%) | 0.405 |
| No | 34 (44.15%) | 33 (42.85%) | 10 (12.98%) | |
| Abortion | | | | |
| Never | 30 (38.46%) | 38 (48.71%) | 10 (12.82%) | 0.089 |
| Once | 11 (57.89%) | 7 (36.84%) | 1 (5.26%) | |
| Twice | 2 (33.33%) | 3 (50.00%) | 1 (16.67%) | |
| 1st degree BC | | | | |
| No | 41 (95.35%) | 46 (95.83%) | 10 (83.33%) | 0.003 |
| 1–2 | 2 (4.65%) | 2 (4.17%) | 0 (0.00%) | |
| > 3 | 0 (0.00%) | 0 (0.00%) | 2 (16.67%) | |
| 2nd degree BC | | | | |
| No | 38 (88.37%) | 41 (85.42%) | 10 (83.33%) | 0.870 |
| 1 | 5 (11.63%) | 7 (14.58%) | 2 (16.67%) | |
| 1st and 2nd degree ovarian cancer | | | | |
| No | 39 (90.70%) | 40 (83.33%) | 11 (91.67%) | 0.511 |
| 1 and 2 | 4 (9.30%) | 8 (16.67%) | 1 (8.33%) | |
| 1st and 2nd degree other cancer | | | | |
| No | 15 (34.88%) | 20 (41.67%) | 3 (25.00%) | 0.219 |
| 1 and 2 | 25 (58.14%) | 19 (39.58%) | 8 (66.67%) | |
| > 3 | 3 (6.98%) | 9 (18.75%) | 1 (8.33%) | |
| Inheritable Disease | | | | |
| No | 30 (69.77%) | 35 (72.92%) | 8 (66.67%) | 0.893 |
| 1–3 | 13 (30.23%) | 13 (27.08%) | 4 (33.33%) | |

duty does not have any significant association with tumor grade to be noted. We also analyzed dietary habits to see if the amount or frequency in the intake of each variables such as, pork, fish, chicken, beef, fruits, vegetables, Sa-um (fermented pork fats) which is a unique Mizo tribal dish, smoked meat, smoked vegetables, salt, oil, and water has impact on tumor grade. Though significant differences in amount and frequencies of intake were observed in the variables, no distinct association was noted between dietary habits and tumor grade (Table 3). Lastly, association between tobacco and

**Table 3** Correlation between tumor grade lifestyle and dietary habits, tobacco, and alcohol habits

| | Grade I (43) | Grade II (48) | Grade III (12) | *p* value |
|---|---|---|---|---|
| Pork | | | | |
| Never | 2 (4.65%) | 9 (18.75%) | 1 (8.33%) | 0.235 |
| Once | 23 (53.49%) | 26 (54.17%) | 5 (41.67%) | |
| Thrice | 12 (27.91%) | 11 (22.92%) | 5 (41.67%) | |
| Everyday | 6 (13.95%) | 2 (4.17%) | 1 (8.33%) | |
| Fish | | | | |
| Never | 0 (0.00%) | 7 (14.58%) | 1 (8.33%) | 0.105 |
| Once | 29 (67.44%) | 30 (62.50%) | 7 (58.33%) | |
| Thrice | 8 (18.60%) | 8 (16.67%) | 4 (33.33%) | |
| Everyday | 6 (13.95%) | 3 (6.25%) | 0 (0.00%) | |
| Chicken | | | | |
| Never | 1 (2.33%) | 4 (8.33%) | 0 (0.00%) | 0.378 |
| Once | 29 (67.44%) | 33 (68.75%) | 7 (58.33%) | |
| Thrice | 7 (16.28%) | 9 (18.75%) | 4 (33.33%) | |
| Everyday | 6 (13.95%) | 2 (4.17%) | 1 (8.33%) | |
| Beef | | | | |
| Never | 3 (6.98%) | 9 (18.75%) | 4 (33.33%) | 0.230 |
| Once | 28 (65.12%) | 29 (60.42%) | 6 (50.00%) | |
| Thrice | 6 (13.95%) | 8 (16.67%) | 1 (8.33%) | |
| Everyday | 6 (13.95%) | 2 (4.17%) | 1 (8.33%) | |
| Fruits | | | | |
| Regular | 15 (34.88%) | 17 (35.42%) | 4 (33.33%) | 0.785 |
| Normal | 27 (62.79%) | 27 (56.25%) | 7 (58.33%) | |
| Never | 1 (2.33%) | 4 (8.33%) | 1 (8.33%) | |
| Vegetables | | | | |
| Regular | 36 (83.72%) | 32 (66.67%) | 8 (66.67%) | 0.243 |
| Normal | 7 (16.28%) | 12 (25.00%) | 3 (25.00%) | |
| Never | 0 (0.00%) | 4 (8.33%) | 1 (8.33%) | |
| Sa-um | | | | |
| Never | 8 (18.60%) | 12 (25.00%) | 4 (33.33%) | 0.796 |
| Normal | 27 (62.79%) | 28 (58.33%) | 7 (58.33%) | |
| Regular | 8 (18.60%) | 8 (16.67%) | 1 (8.33%) | |
| Smoked meat | | | | |
| Never | 5 (11.63%) | 7 (14.58%) | 2 (16.67%) | 0.773 |
| Normal | 35 (81.40%) | 36 (75.00%) | 10 (83.33%) | |
| Regular | 3 (6.98%) | 5 (10.42%) | 0 (0.00%) | |
| Smoked veg | | | | |
| Never | 11 (25.58%) | 14 (29.17%) | 2 (16.67%) | 0.733 |
| Normal | 29 (67.44%) | 31 (64.58%) | 8 (66.67%) | |
| Regular | 3 (6.98%) | 3 (6.25%) | 2 (16.67%) | |
| Salt intake | | | | |
| Less | 15 (34.88%) | 15 (31.25%) | 3 (25.00%) | 0.800 |
| Heavy | 28 (65.12%) | 33 (68.75%) | 9 (75.00%) | |
| Oil intake | | | | |
| Less | 14 (32.56%) | 15 (31.25%) | 3 (25.00%) | 0.882 |
| Heavy | 29 (67.44%) | 33 (68.75%) | 9 (75.00%) | |
| Water intake | | | | |
| > 2 L | 6 (13.95%) | 7 (14.58%) | 2 (16.67%) | 0.985 |
| 1–2 L | 9 (20.93%) | 14 (29.17%) | 3 (25.00%) | |
| 500–1 L | 23 (53.49%) | 22 (45.83%) | 6 (50.00%) | |
| 1–2 glass | 5 (11.63%) | 5 (10.42%) | 1 (8.33%) | |
| Exercise | | | | |
| Never | 20 (46.51%) | 27 (56.25%) | 3 (25.00%) | 0.153 |
| 3–4 times | 10 (23.26%) | 8 (16.67%) | 6 (50.00%) | |
| Everyday | 13 (30.23%) | 13 (27.08%) | 3 (25.00%) | |
| Sleeping hours | | | | |
| 1–3 | 2 (4.65%) | 4 (8.33%) | 1 (8.33%) | 0.151 |
| 3–6 | 10 (23.26%) | 18 (37.50%) | 5 (41.67%) | |
| 6–8 | 27 (62.79%) | 21 (43.75%) | 5 (41.67%) | |
| 8–10 | 4 (9.30%) | 5 (10.42%) | 1 (8.33%) | |
| Night duty | | | | |
| Yes | 9 (20.93%) | 6 (12.50%) | 3 (25.00%) | 0.438 |
| No | 34 (79.07%) | 42 (87.5%) | 9 (75.00%) | |
| Betel nut/pan | | | | |
| Never | 8 (18.60%) | 10 (20.83%) | 5 (41.67%) | 0.278 |
| Less | 14 (32.56%) | 7 (14.58%) | 2 (16.67%) | |
| Normal | 10 (23.26%) | 16 (33.33%) | 3 (25.00%) | |
| Heavy | 11 (25.58%) | 15 (31.25%) | 2 (16.67%) | |
| Gutkha | | | | |
| Never | 39 (90.70%) | 44 (91.67%) | 12 (100.00%) | 0.978 |
| Less | 2 (4.65%) | 2 (4.17%) | 0 (0.00%) | |
| Normal | 1 (2.33%) | 1 (2.08%) | 0 (0.00%) | |
| Heavy | 1 (2.33%) | 1 (2.08%) | 0 (0.00%) | |
| Sahdah | | | | |
| Never | 13 (30.23%) | 11 (22.92%) | 6 (50.00%) | 0.111 |
| Less | 11 (25.58%) | 4 (8.33%) | 1 (8.33%) | |
| Normal | 6 (13.95%) | 12 (25.00%) | 2 (16.67%) | |
| Heavy | 13 (30.23%) | 21 (43.75%) | 3 (25.00%) | |
| Khaini | | | | |
| Never | 33 (76.74%) | 32 (66.67%) | 10 (83.33%) | 0.327 |
| Less | 4 (9.30%) | 7 (14.58%) | 0 (0.00%) | |
| Normal | 4 (9.30%) | 2 (4.17%) | 0 (0.00%) | |
| Heavy | 2 (4.65%) | 7 (14.58%) | 2 (16.67%) | |
| Tuibur | | | | |
| Never | 32 (74.42%) | 30 (62.50%) | 6 (50.00%) | 0.593 |
| Less | 4 (9.30%) | 10 (20.83%) | 3 (25.00%) | |
| Normal | 5 (11.63%) | 4 (8.33%) | 2 (16.67%) | |
| Heavy | 2 (4.65%) | 4 (8.33%) | 1 (8.33%) | |
| Cigarette | | | | |
| Never | 24 (55.81%) | 34 (70.83%) | 8 (66.67%) | 0.562 |
| Less | 6 (13.95%) | 6 (12.50%) | 1 (8.33%) | |
| Normal | 7 (16.28%) | 2 (4.17%) | 2 (16.67%) | |
| Heavy | 6 (13.95%) | 6 (12.50%) | 1 (8.33%) | |
| Alcohol | | | | |
| Never | 41 (95.35%) | 48 (100.00%) | 12 (100.00%) | 0.241 |
| Less | 2 (4.65%) | 0 (0.00%) | 0 (0.00%) | |

alcohol consumption with tumor grade was also analyzed, and the variables did not show significant results in determining the tumor grade.

Since no association was observed, analysis was carried out again by confining our search among age group 40–55 years at the time of BC diagnosis. However, we did not observe any significant association between tumor grades with different lifestyle, dietary, tobacco, and alcohol habits (Supplementary Table 4).

## Discussion

In India, BC patients tend to be younger, tumors are often larger when first diagnosed, and are of high grade as compared with Western countries (Kaul et al. 2011). The mean age at diagnosis is about 48 years in Mizoram and is similar to several studies conducted in Africa, Ivory Coast, and Middle East, while in developed countries like the USA and Western Europe, it is 63 years and around 51 years in Iran where BC commonly occurs at the advanced age or at the post-menopausal period (Mirunalini et al. 2010; Pourzand et al. 2011; Effi et al. 2017).

To help patients in decision-making and provide the best possible treatment, current BC management depends on the available clinical and pathological predictive and prognostic factors. Histological tumor grade is a well-established prognostic factor as it evaluates the morphology of tumor biological characteristics and can provide critical reports that are linked to BC clinical behavior (Rakha et al. 2010). Elston and Ellis (1991) had suggested that the Nottingham modification of Bloom and Richardson which has been the most commonly used and widely accepted method for histological grading system should be used as a standard prognostic factor for all BC patients, since there is a strong evidence that tumor grade associates well in the biochemical and kinetic indicator of differentiation.

Lymph node status and Nottingham grading system have been shown to be equivalently important prognostic factor of BC (Rakha et al. 2010). It also helped in determining cancer staging and treatment options. Even though there is a limited existing literature, correlating direct association between tumor grades with lymph node involvement, there are studies correlating tumor grade with the number of lymph node involved and found to be significantly correlated. With the increase in the number of positive axillary lymph nodes, survival rate decreases and relapse rate increases (Siadati et al. 2015; Davis et al. 1986). It had also been shown that a lymph node negative patient had better survival results than a patient with positive lymph node; they also added that a patient with more than 10 positive lymph nodes presented the poorest survival outcomes regardless of the tumor size (Wang et al. 2016). Falling in line, our results show association of tumor grade with lymph node even though we cannot conclude to what extent they are associated.

For routine practice, the three common prognostic factors used in early-stage BC are histological tumor grade, lymph node status, and tumor size (Rakha et al. 2010). In our study, we have found an association between tumor grades with tumor size among age 40–55 years. Henson et al. (1991) have found that, even if there is a positive lymph node invasion, patients having grade 1 tumors with < 2 cm in size showed an exceptionally good prognosis (99% 5-year survival). A 20-year study on Swedish Two-County screen-detected BC patients has demonstrated that lymph node status and tumor grade and size at diagnosis time show a lasting influence on their overall survival (Warwick et al. 2004).

In our study, significant correlation was observed between tumor grade and status of ER, PR, and HER2/neu that is similar to the results of Siadati et al. (2015). Patients with ER+ (54.36%) are superior to those of PR+ (49.51%), quiet similar to Haiti cohort (51.8%) and is also quite comparable with the findings reported by different authors from the USA, Ivory Coast, Europe, and Africa (Effi et al. 2017; De Gennaro et al. 2018). Interestingly, contrasting to our findings is the study from the south Indian population with ER− patients much higher (61.7%) than the ER+ (Thiygarajan et al. 2015). We have found status of ER to have an association with tumor grade, and our results supports the previous findings from Western Chinese populations (Zheng et al. 2018). In our studies, with the increasing grade, ER+ was observed to be decreasing. Grade III tumor was found to have less ER+ as compared with grade I and grade II tumor. This result is in concordance with the study performed by Atif et al. (2018).

Additionally, while our findings on hormone receptors status, ER+/PR+ (46.60%) and ER−/PR− (42.71%) being the most frequent subtypes are in corroborating with the results of many studies; there are also studies reporting contrasting results especially for ER+/PR+. An inverse relationship between the hormone receptor status and tumor grade has been observed. The greater the tumor grade, the lesser the ER+/PR+ found. Similar results have been reported (Atif et al. 2018; Effi et al. 2017; Zheng et al. 2018). Our results suggest the treatment options, such that ER+/PR+ patients could undergo hormonal treatment while ER−/PR− patients should benefit from chemotherapy.

In our studies, the frequency of HER2/neu+ is 49.51% and is very much similar to many international and local studies. Thiygarajan et al. (2015) observed an almost similar result on HER2/neu+ (43.3%) among the south Indian population. However, in contrast to our results, HER2/neu+ was observed at a relatively low percentage (9%) in Maharastra, India (George et al. 2018). This suggested that there are variations on HER2/neu reactivity among the Indian state, that is, between the western and northeastern part of India and also explains the role of difference in ethnicity.

HER2/neu is an epidermal growth factor present on the cell surfaces that transmit growth signals to the cell nucleus. Over

expression of HER2/neu oncoprotein is associated with poor prognosis. In our study, significant correlation has been observed between tumor grades with status of HER-2/neu oncoprotein. Several other studies have found a significant correlation between histological grade III and HER-2/neu expression (Siadati et al. 2015). Similar to our finding, it has also been reported that HER2/neu+ is more common in high-grade tumor (Zheng et al. 2018). However, in contrast to our results, Thiygarajan et al. (2015) observed no correlation between tumor grades and HER2/neu+.

It is well known that triple negative BC is associated with early age at diagnosis and is characterized by poor prognosis, high invasiveness, and recurrence within the first 3 years, as well as a higher 5-year mortality rate than other BC subtypes (Wang et al. 2016; De Gennaro et al. 2018). It is also known that the prevalence of triple negative BC differs among different ethnic groups. Several studies from different countries have reported variations in hormone receptors and Her2/neu status in BC patients. Sandhu et al. (2016) have reported the prevalence of TNBC in India (which compares all four region, the north, east, west, and south) to be considerably higher (31%) compared with Western populations. But contrastingly in our studies, 17.47% of BCs have triple negative subtype, which is quite similar to the findings on Southeastern Turkey population (Mirunalini et al. 2010; Sandhu et al. 2016). From our studies, we have also found that ER/PR/HER2 status is significantly associated in determining the grade of tumor. There are other studies that separately correlate hormone receptor status and HER2/neu status with grade of tumors. They have observed the ER/PR expressions to have a significant correlation with tumor grade I and II and have also found that the expression of HER2/neu to be associated with tumor grade III (Siadati et al. 2015).

Family history is among the most important risk factor for the onset of BC with a relative risk associated with the number of affected individuals, their age at diagnosis, and the degree of relationship. It has been reported that 10–30% of women with BC have at least one or more relative with the same disease (Tazzite et al. 2013). Existing literatures that correlate breast tumor characteristics with family history of BC had given a contradicting finding. In search for the association between family history of BC and severity of the disease at the time of diagnosis, Melvin et al. (2016) examined different tumor grades and found that family history does not appear to have an association with BC severity at the time of diagnosis, nor BC-mortality. However, Slattery et al. (1993) have reported that BC patients younger than 50 years of age with a family history have a significant greater risk of poorer survival when compared with those women without a family history, with a relative risk of 1.54 for women with first degree relatives previously diagnosed with BC.

In our study, 20 (19.41%) patients have a positive family history of BC. Out of which, 6 patients (5.82%) have first degree relative diagnosed with BC. The results of our study is quite similar to the finding of Slattery et al. (1993) because the mean age of a patient with a first degree relative is much younger, which is 49 years and also have statistical significant association with the tumor grade while no association was observed with patients having a second degree relative.

Even though, our studies do not show any significant correlation between tumor grade and different epidemiological factors studied; it is well known that many of the factors included in the study had been consistently reported to increase the risk of causing BC. Several studies on reproductive factors like age at menarche, age at menopause, age at first delivery, parity, lack of breast feeding, duration of lactation, age at menopause, and use of menopausal hormone therapy and oral contraceptive pills have been reported to increase BC risk. Lack of breast feeding has also been shown to have a positive association with BC (Sweeney et al. 2008; Lodha et al. 2011; Anstey et al. 2017).

With societal development, habits relating to lifestyles such as pattern of sleep and night time/shift work have drawn the attention of researchers, as sleep might affect the levels of several circulating hormones, which include melatonin, insulin, growth hormone, prolactin, glucose, and cortisol that play an important role in the development of many diseases, including BC (Lu et al. 2017). A large cohort study analyzing 73,615 post-menopausal women had described a 25% lower risk for BC in women with regular physical exercise than those inactive women (Hildebrand et al. 2013).

The Nurses' Health Study and a study conducted in NIH-AARP Diet and Health and also several other studies had observed heavy intake of animal fats, red meat, smoked, and processed meat and lower intake of fruits, vegetables, and total dietary fiber to have significant association with BC risk (Haraldsdottir et al. 2018; Kim et al. 2017). Taken together, our results suggest the important roles played by several clinico-pathological and epidemiological factors in determining the histological tumor grade among Mizo population.

## Conclusion

Our results showed significant correlation between family history with first degree BC, status of lymph node, hormone receptors, and HER2/neu oncoprotein in determining the histological grade of breast tumor. The report of our study is the first time to target and focus on the possible role of biomarkers for diagnosis in terms of family history and tumor grade in BC. To conclude, our results further support the previous findings that observed immunohistochemistry test, and sentinel lymph node biopsy results could help patients in making decision and for the pathologist to provide the best treatment option. We also suggest the significance of considering a complete family history of the patients in relation to BC, which

might be an important factor in determining histological tumor grade. However, a study with a bigger sample size is needed to establish the role of family history of BC. Furthermore, our results also demonstrate the important association between tumor grades with tumor size, which is a time dependent prognostic factors. The present study highlights the importance of the factors studied and might help prevent misdiagnosis in rural area like Mizoram and provide challenges for all round development in diagnostic purposes to the patients for a best treatment option. More in-depth studies are needed in this region, where incidence of BC cases is more than 500, out of which, only 426 patients could be reached within 5 years (2013–2017) of sample collection in a population of just around 12.85 lakhs in 2018.

**Acknowledgments** The authors would like to acknowledge the Department of Biotechnology, New Delhi, for the Advanced State Level Biotech Hub and Bioinformatics Infrastructural Facility, Mizoram University, for the completion of this work.

## Compliance with ethical standards

**Competing interests** The authors declare that they have no competing interests.

**Ethics approval** The ethical approval has been obtained from ethical committees of Civil Hospital, Aizawl (B.12018/1/13-CH(A)/IEC/33 dtd.15/10/2014) and Institutional Human Ethical Committee, Mizoram University.

**Informed consent** Informed consent was obtained from all individual participants included in the study.

**Abbreviations** BC, breast cancer; ER, estrogen; PR, progesterone; HER2/neu, HER2/neu oncoprotein; +, positive; −, negative

## References

Anstey EH, Shoemaker ML, Barrera CM, O'Neil ME, Verma AB, Holman DM (2017) Breastfeeding and breast cancer risk reduction: implications for black mothers. Am J Prev Med 53:S40–S46. https://doi.org/10.1016/j.amepre.2017.04.024

Atif N, Khalid M, Chughtai O, Saad A, Rashid M, Chughtai AS (2018) Role of immunohistochemical markers in breast cancer and their correlation with grade of tumour, our experience. Int Clin Pathol J 6(3):141–145. https://doi.org/10.15406/icpjl.2018.06.00175

Davis BW, Gelber RD, Goldhirsch A, Hartmann WH, Locher GW, Reed R, Golouh R, Save-Doderbergh J, Hollowway L, Russel I, Rudenstam CM (1986) Prognostic significance of tumor grade in clinical trials of adjuvant therapy for breast cancer with axillary lymph node metastasis. Cancer 58:2662–2670

De Gennaro V Jr, Jiwani F, Patberg E, Gibbs M, Libby R, Gabriel D, Heldermon CD, Daily K, Bernard J Jr (2018) Epidemiological, clinical, and histopathological features of breast cancer in Haiti. Journal of Global Oncology 4(4):1–9. https://doi.org/10.1200/jgo.17.00135

Effi AB, Aman NA, Koui BS, Koffi KD, Traore ZC, Kouyate M (2017) Immunohistochemical determination of estrogen and progesterone receptors in breast cancer: relationship with clinicopathologic factors in 302 patients in Ivory Coast. BMC Cancer 17:115. https://doi.org/10.1186/s12885-017-3105-z

Elston CW, Ellis IO (1991) Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19: 403–410

George GA, Bembde AS, Borde N, Bhale C (2018) Correlation of histologic grade of breast carcinoma with hormone receptor status in a tertiary care hospital. Int J Res Med Sci 6(4):1331–1336. https://doi.org/10.18203/2320-6012.ijrms20181292

Haraldsdottir A, Torfadottir JE, Valdimarsdottir UA, Adami HO, Aspelund T, Tryggvadottir L, Thordardottir M et al (2018) Dietary habits in adolescence and midlife and risk of breast cancer in older women. PLoS ONE 13(5):e0198017. https://doi.org/10.1371/journal.pone.0198017

Henson DE, Ries L, Freedman LS, Carriaga M (1991) Relationship among outcome, stage of disease, and histologic grade for 22,616 cases of breast cancer. The basis for a prognostic index. Cancer 68: 2142–2149

Hildebrand JS, Gapstur SM, Campbell PT, Gaudet MM, Patel AV (2013) Recreational physical activity and leisure-time sitting in relation to postmenopausal breast cancer risk. Cancer Epidemiol Biomarkers Prev 22(10):1906–1912. https://doi.org/10.1158/1055-9965.EPI-13-0407

Kaul R, Sharma J, Satinder S, Minhas, Mardi K (2011) Hormone receptor status of breast cancer in the Himalayan region of Northern India. Indian J Surg 73(1):9–12. https://doi.org/10.1007/s12262-010-0121-5

Kaur M, Kaur H, Manjari M, Rai V, Kaur K (2016) Assessment of clinical parameters and histopathological grading of breast cancer. International Journal of Contemporary Medical Research 3(10): 2938–2941

Kim JH, Lee J, Jung SY, Kim J (2017) Dietary factors and female breast cancer risk: a prospective cohort study. Nutrient 9:1331. https://doi.org/10.3390/nu9121331

Kuzhan A, Adli M, Alkis HE, Caglayan D (2013) Hormone receptor and HER2 status in patients with breast cancer by races in southeastern Turkey. JBUON 18(3):619–622

Lodha R, Joshi A, Paul D, Lodha KM, Nahar N, Shrivastava A, Bhagat VK, Nandeshwar S (2011) Association between reproductive factors and breast cancer in an urban set up at central India: a case-control study. Indian J Cancer 48:3. https://doi.org/10.4103/0019-509X.84928

Lu C, Sun H, Huang J, Yin S, Hou W, Zhang J, Wang Y, Xu Y, Xu H (2017) Long-term sleep duration as a risk factor for breast cancer: evidence from a systematic review and dose-response meta-analysis. Biomed Res Int. https://doi.org/10.1155/2017/4845059

Madathil S, Kumar NS, Zodinpuii D, Muthukumaran RB, Lalmuanpuii R, Nicolau B (2018) Tuibur: tobacco in a bottle- commercial production of tobacco smoke- saturated aqueous concentrate. Addiction 113(3):577–580. https://doi.org/10.1111/add.14117

Melvin JC, Wulaningsih W, Hana Z, Purushotham AD, Pinder SE, Fentiman I, Gillett C, Mera A, Holmberg L, Hemelrijck MV (2016) Family history of breast cancer and its association with disease severity and mortality. Cancer Med 5(5):942–949. https://doi.org/10.1002/cam4.648

Mirunalini S, Dhamodharan G, Karthishwaran K, Sugunadevi G (2010) Correlation of estrogen and progesterone receptors status with the grade and type of breast cancer. Res J Agric & Biol Sci 6(3):199–203

Mizoram Population Census (2018) Literacy rate. https://www.census2011.co.in/census/state/mizoram.html. Accessed 16 April 2019

National Cancer Registry Programme (NCRP) 2014. http://www.ncdirindia.org/ncrp/ALL_NCRP_REPORTS/

PBCR_REPORT_2012_2014/ALL_CONTENT/Printed_Version.htm. Accessedr 16 April 2019

Perez-Solis MA, Maya-Nuñez G, Casas-González P, Olivares A, Aguilar-Rojas A (2016) Effects of the lifestyle habits in breast cancer transcriptional regulation. Cancer Cell Int 16:7. https://doi.org/10.1186/s12935-016-0284-7

Pourzand A, Fakhree MBA, Hashemzadeh S, Halimi M, Daryani A (2011) A hormone receptor status in breast cancer and its relation to age and other prognostic factors. Breast Cancer: Basic and Clinical Research 5:87–92. https://doi.org/10.4137/BCBCR.S7199

Rakha EA, Reis-Filho JS, Baehner F, Dabbs DJ, Decker T, Eusebi V, Fox SB, Shu Ichihara S, Jacquemier J, Lakhani SR, Palacios J, Richardson AL, Schnitt SJ, Schmitt FC, Tan PH, Tse GM, Badve S, Ellis IO (2010) Breast cancer prognostic classification in the molecular era: the role of histological grade. Breast Cancer Res 12:207. https://doi.org/10.1186/bcr2607

Sandhu GS, Erqou S, Patterson H, Mathew A (2016) Prevalence of triple-negative breast cancer in India: systematic review and meta-analysis. Journal of Global Oncology 2(6):412–421. https://doi.org/10.1200/JGO.2016.005397

Siadati S, Sharbatdaran M, Nikbakhsh N, Ghaemian N (2015) Correlation of ER, PR and HER-2/Neu with other prognostic factors in infiltrating ductal carcinoma of breast. Iran J Pathology 10(3):221–226

Slattery ML, Berry TD, Kerber RA (1993) Is survival among women diagnosed with breast cancer influenced by family history of breast cancer? Epidemiology 4(6):543–548

Sweeney C, Baumgartner KB, Byers T, Giuliano AR, Herrick JS, Murtaugh MA, Slattery ML (2008) Reproductive history in relation to breast cancer risk among Hispanic and non-Hispanic white women. Cancer Causes Control 19(4):391–401. https://doi.org/10.1007/s10552-007-9098-1

Tazzite A, Jouhadi H, Saiss K, Benider A, Nadifi S (2013) Relationship between family history of breast cancer and clinicopathological features in Moroccan patients. Ethiop J Health Sci 23:2

Thiygarajan M, Navrathan N, Mohanapriya T, Kumar A, Singh B (2015) Correlation between estrogen receptor, progesterone receptor, HER-2/neu status and other prognostic factors in carcinoma breast in Indian population. Int Surg J 2(4):515–522. https://doi.org/10.18203/2349-2902.isj20151000

Tobacco, Mizoram State Tobacco Control Programme (MSTCP), Health & Family Welfare Department, Government of Mizoram https://health.mizoram.gov.in/page/tobacco. Accessed 16 April 2019

Wang XX, Jiang YZ, Li JJ, Song CG, Shao ZM (2016) Effect of nodal status on clinical outcomes of triple-negative breast cancer: a population-based study using the SEER 18 database. Oncotarget 7:29

Warwick J, Tabar L, Vitak B, Duffy SW (2004) Time-dependent effects on survival in breast carcinoma: results of 20 years of follow-up from the Swedish Two-County Study. Cancer 100:1331–1336

Zheng K, Tan JX, Li F, Li HY, Zeng XH et al (2018) Clinicopathologic factors related to the histological tumor grade of breast cancer in Western China: an epidemiological multicenter study of 8619 female patients. Translational Oncology 11(4):1023–1033. https://doi.org/10.1016/j.tranon.2018.06.005

# Environmental Science and Pollution Research

ISSN: 0944-1344 (Print) 1614-7499 (Online)


## Description

Environmental Science and Pollution Research (ESPR) serves the international community in all areas of Environmental Science and related subjects with emphasis on chemical compounds. It reports from a broad interdisciplinary outlook. Apart from the strictly scientific contributions as research articles (short and full papers) and reviews, ESPR publishes: news & views from research and technology, legislation and regulation, h … show all

Browse Volumes & Issues

| Impact Factor | Available |
|---|---|
| 2.914 | 1994 - 2019 |
| **Volumes** | **Issues** |
| 26 | 304 |

## Latest Articles

# Letter to the Editor

## TUIBUR: TOBACCO IN A BOTTLE—COMMERCIAL PRODUCTION OF TOBACCO SMOKE-SATURATED AQUEOUS CONCENTRATE

Certain forms of tobacco that are relatively unknown in the scientific community have been used widely for centuries. One such form is tuibur [1,2], a tobacco smoke-saturated aqueous concentrate prevalent in north-eastern India [3].

The highest prevalence of tobacco consumption in India occurs in the northeastern province of Mizoram [4,5], where the Aizawl District also leads the incidence rates of all tobacco-related cancers in the country [6]. Yet references to tuibur use are rare and restricted mainly to local publications [3,7–12].

This commentary aims to raise awareness and stimulate research on this tobacco product about which little is known. We describe its production and distribution with the support of photographs to contextualize this knowledge for readers.

Historically, women smoked tobacco in waterpipes made of bamboo (Fig. 1), and would offer the leftover tobacco-infused water from the pipe reservoir to their husband [10]. Users keep a small amount (5–10 ml) of liquid in the mouth for many minutes several times throughout the day; some may also swallow a portion of the liquid [13].

Increased demand for the product moved production from households to makeshift sheds located near natural water sources (Fig. 2). Indeed, concern over its environmental and health impact may help to explain why tuibur is produced in small batches in artisanal facilities [14–16]. The production of a fresh batch of tuibur starts early in the morning. First, water is alkalized by being filtered through tobacco ash from the previous batch (Supporting information, Fig. S1). Alkaline pH maintains an unprotonated form of nicotine that is absorbed easily through the oral mucosa [1]. This solution is then poured into the lower chamber of a device (Fig. 3) with a rudimentary water educator (aspirator). A pipe is inserted into this chamber with one end just below the water level, and sealed with mud (Fig. 3). The upper half of the device, the combustion chamber, is connected to the lower chamber through a filtered opening at the bottom.

Dried tobacco stalk is placed in the combustion chamber and set to smoulder (Supporting information, Fig. S2). Subsequently, this chamber is sealed with wet tobacco ash. The pressure from a stream of water creates a suction in the smaller pipes, which ensures that tobacco smoke bubbles through the alkaline solution. The process

**Figure 1** Traditional waterpipe used for smoking tobacco (Mizoram) [Colour figure can be viewed at wileyonlinelibrary.com]

**Figure 2** Tuibur production facility constructed on a cliff near a natural stream, outskirts of Aizawl city, Mizoram [Colour figure can be viewed at wileyonlinelibrary.com]

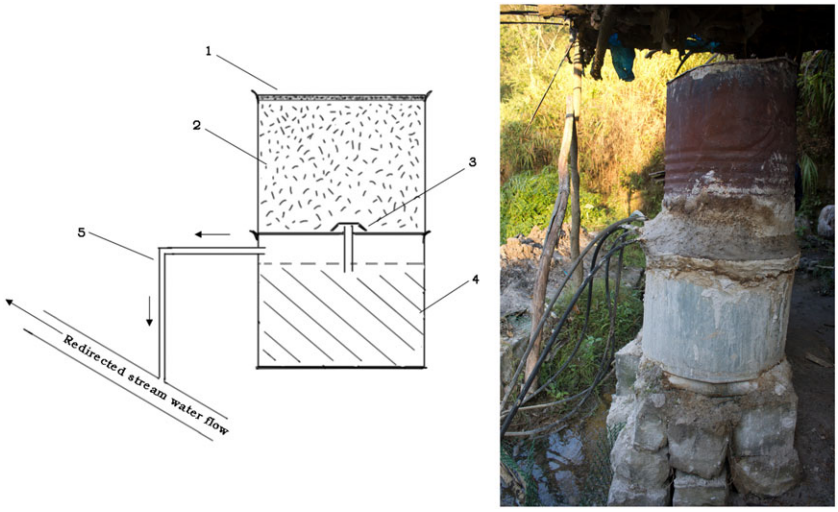



**Figure 3** Illustration and photograph of the device used to produce tuibur. (1) Tobacco ash; (2) dried tobacco stalk; (3) filter to prevent tobacco from falling into the lower compartment; (4) alkaline feedstock; (5) small pipe to inhale the tobacco smoke [Colour figure can be viewed at wileyonlinelibrary.com]

is complete when the tobacco stalk has burned completely, which takes anywhere from 12 hours to 3 days, depending upon the amount of tobacco stalk used. 'Crude' tuibur from the lower chamber is then pipetted out and filtered again through tobacco ash (Supporting information, Figure S3). This 'commercial' grade tuibur is sold in bulk to distributors, who bottle and sell it in local markets with a statutory pictorial warning.

Tuibur consumption is highly addictive, and deeply entrenched in the culture of the region. In addition, the rudimentary character of tuibur production has potential environmental consequences, which may compound health impacts. At a time when tobacco control is a global priority, we argue that this deleterious habit merits more research attention.

## Declaration of interests

None.

## Acknowledgements

We are grateful to the producers of Tuibur who welcomed our visit to their production site. We also thank Drs G.

Castonguay and A. Levine for providing helpful comments on the manuscript. B. Nicolau holds a Canada Research Chair in Life Course Oral Epidemiology and received a Shastri Scholar Travel Subsidy Grant to support her visit to University of Mizoram.

**Keywords** Habits, mouth, nicotine, northeastern India, smokeless tobacco, tuibur.

SREENATH MADATHIL[1], NACHIMUTHU SENTHIL KUMAR[2], DORIS ZODINPUII[2], RAJENDRA B. MUTHUKUMARAN[3], REBECCA LALMUANPUII[3] & BELINDA NICOLAU[1]

*Division of Oral Health and Society, Faculty of Dentistry, McGill University, Montreal, Canada,[1] Department of Biotechnology, Mizoram University, Aizawl, Mizoram, India[2] and Department of Chemistry, Mizoram University, Aizawl, Mizoram, India[3]*

*E-mail: belinda.nicolau@mcgill.ca*

## References

1. Muthukumaran R. B. Tobacco-specific nitrosamines in tuibur. *Sci Vis*16 2016; **16**: 95–105.
2. Mahanta J., Chetia M., Hazarika N. C., Narain K., Sharma S. K. Toxicity of tuibur, a unique form of tobacco smoke extract used in Mizoram, India. *Curr Sci* 1998; **75**: 381–4.
3. Sinha D. N., Gupta P. C., Pednekar M. Tobacco water: a special form of tobacco use in the Mizoram and Manipur states of India. *Natl Med J India* 2004; **17**: 245–7.
4. Census of India 2001. 2001; 27–8.
5. International Institute for Population Sciences (IIPS). Global Adult Tobacco Survey India 2009–2010 [internet]. GATS. Deonar, Mumbai: IIPS; 2010. Available at: http://www.searo.who.int/tobacco/documents/2010-pub2.pdf (accessed 18 December 2017) (Archived at http://www.webcitation.org/6vo26dYi4).
6. National Cancer Registry Programme. Comparison of cancer incidence and patterns of all population based cancer registries in India [internet]. Three-year report of population based cancer: 2012–2014. 2015. Available at: http://ncrpindia.org/ALL_NCRP_REPORTS/PBCR_REPORT_2012_2014/ALL_CONTENT/Printed_Version.htm (accessed 18 December 2017) (Archived at http://www.webcitation.org/6vo029nm5).
7. Phukan R. K., Zomawia E., Narain K., Hazarika N. C., Mahanta J. Tobacco use and stomach cancer in Mizoram, India. *Cancer Epidemiol Biomarkers Prev* 2005; **14**: 1892–6.
8. Sinha D. N., Gupta P. C., Pednekar M. S. Use of tobacco products as dentrifice among adolescents in India: questionnaire survey. *BMJ* 2004; **328**: 323–4.
9. Sinha D. N., Gupta P. C., Pednekar M. S. Tobacco use among students in the eight northeastern states of India. *Indian J Cancer* 2003; **40**: 43–59.
10. Gupta P., Arora M., Sinha D., Asma S., Parascondola M., editors. *Smokeless Tobacco and Public Health in India*. Ministry of Health and Family Welfare, Government of India; 2016. Available at: https://www.mohfw.gov.in/sites/default/files/Final%20Version%20of%20SLT%20Monograph.pdf (accessed 18 December 2017) (Archived at http://www.webcitation.org/6voHMoOwE).
11. Lalmuanpuii R., Muthukumaran R. B. Vibrational spectroscopic characterization of tuibur: an indigenous nicotine delivery medium. *SciVis16* 2016; **16**: 53–8.
12. Sinha D., Gupta P., Subramoney S. Carcinogens in tobacco water. *Natl Med J India* 2006; **19**: 110–1.
13. International Agency for Research on Cancer (IARC). IARC monographs on the evaluation of carcinogenic risks to humans—smokeless tobacco and some tobacco-specific N-nitrosamines. *IARC Monograph* 2007; **89**: 223–76.
14. Sanga Z. Cancer scare for tuibur users—environmentalists warn of river pollution in Mizoram. The Telegraph—India [internet]. March 20 2014. Available at: https://www.telegraphindia.com/1140320/jsp/northeast/story_18097542.jsp (accessed 3 August 2017) (Archived at http://www.webcitation.org/6sR3vePf3).
15. Sanga Z. Closure of Aizwal tobacco factories directed on health ground. Assam Tribune 11 February 11 2017.
16. Lalrinpuii E. TUIBUR FACTORY CHUNGCHANG SAWI [internet]. Press release. Aizawl; 2013. Available at: url:https://dipr.mizoram.gov.in/press-release/tuibur-factory-chungchang-sawi (accessed 3 August 2017) (Archive at http://www.webcitation.org/6sRLXK3tz).

## Supporting Information

Additional Supporting Information may be found online in the supporting information tab for this article.

**Figure S1** Preparation of feedstock for tuibur. S1a: Filtering water through tobacco ash to make it alkaline; S1b: filling the alkaline feedstock to lower compartment; S1c: inlet pipe for tobacco smoke to lower chamber.
**Figure S2** S2a and S2b: Dried tobacco stalk loaded into the upper chamber; S2c: burning of tobacco stalk in the upper chamber.
**Figure S3** S3a: Pipetting out of tuibur from lower chamber; S3b: filling containers with tuibur for transportation; S3c: tuibur sold in local market.

**PARTICULARS OF THE CANDIDATE**

| | | |
|---|---|---|
| NAME OF THE CANDIDATE | : | Doris Zodinpuii |
| DEGREE | : | Ph.D |
| DEPARTMENT | : | Biotechnology |
| TITLE OF THE THESIS | : | Characterization of clinically significant mutations associated with Breast Cancer in Mizo population |
| DATE OF ADMISSION | : | 11.08.2015 |
| APPROVAL OF RESEARCH PROPOSAL | | |
| 1. BOARD OF STUDIES | : | 7.04.2016 |
| 2. SCHOOL BOARD | : | 22.04.2016 |
| REGISTRATION NO. & DATE | : | Mzu/Ph.D/939 of 15.11.2016 |
| EXTENSION (IF ANY) | : | NA |

(Dr. Th. Robert Singh)

Head

Department of Biotechnology

**Questionnaire for Functional Characterization of clinically significant *h-BRCA1***

**Gene mutation 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) :

Hardly / 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?**
(Zu I in thin em?I in thin cuan eng anga zing/tam nge?) :

(Local / Branded / Both) 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?**
**(I hnute khawi lamah nge bawk an hmuhchhuah?)**
**Right ( ) Left ( ) Both sides ( )**

**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)**

# ABSTRACT

# CHARACTERIZATION OF CLINICALLY SIGNIFICANT MUTATIONS ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

**DORIS ZODINPUII**

**DEPARTMENT OF BIOTECHNOLOGY**
**MIZORAM UNIVERISTY**

# ABSTRACT

**Characterization of clinically significant mutations associated with Breast Cancer in Mizo population**

**By**

**DORIS ZODINPUII**

**Department of Biotechnology**

**Submitted**

**In partial fulfillment of therequirement of the Degree of Doctor of Philosophy in Biotechnology of Mizoram Univeristy, Aizawl**

***Abstract***

**Introduction**

Breast Cancer (BC) is a complex multifactorial disease regulated by the association of various genetic, hormonal and environmental factors including diet, lifestyle and reproductive history. Globally, BC is the top cancer and causes of death among women. BC related death of around 327,000 has been estimated every year. In India, almost 100,000 cases of BC have been diagnosed each year, and an increased to 131,000 cases is expected by the year 2020. In the North Eastern part of India, the prevalence of BC is considerably higher then the rest of India probably due to the tobacco exposure leading to genotoxic stress. Mizoram, located in the Southern corner of Northeast India, also shows an alarming trend in cancer. Latest report of National Cancer Registry Program (NCRP) revealed consistently high incidence rate of cervix uteri (15.9 %), lung (15.6 %), and BC (13.0 %) among the Mizo women populace. BC incidence rate is believed to be already double by 2018 since the recent NCRP is not yet available. Variations in the incidence and mortality rates of BC between races or ethnicity of different parts of the world suggests that the known risk factors for BC may differ depending on the demographic and environmental factors and could not be explained by genetics alone. Several factors that have been reported to help in BC progression are reproductive factors, lifestyle and dietary habits, tobacco and alcohol intake, and family history of breast cancer. With the evolving sequencing technologies, number of BC susceptibility genes has been reported. *BRCA1* and *BRCA2* are the two most crucial BC predisposing genes that have a major contribution in the development of inherited BC. Germline

mutation carriers in these two high penetrant genes increases the lifetime risk of developing breast as well as ovarian cancer by 80% and are also known to be associated with the disease onset at an early age. Several studies performing a comprehensive multi-gene panel testing for BC has unraveled the remaining responsible genes other than *BRCA1/2*. A large number of gene mutations (~36-61%) have been observed in BC patients including those among the high (*PTEN, CDH1, TP53, STK11 and NF1*), moderate (*NBN* and related genes, *CHEK2, PALB2, ATM*) and low penetrance gene (*PMS2, MEN1, MLH1, MSH6, PPM1D, MSH2* and so on) indicating that multiple gene panel testing serves as a more efficient tool for genetic cancer risk assessment for different sub-types of BC. Mitochondria play an important role in BC pathogenesis. They produce energy through aerobic respiration, making mitochondria the main source and target of intracellular reactive oxygen species (ROS). Mutation rates are much higher as compared to nuclear DNA most likely because of less efficiency in damage DNA repair mechanism. Multiple studies have also found associations between cancer development and somatic mtDNA mutations, also hereditary mtDNA polymorphism have been found to contribute in cancer progression. From the recent years, mtDNA have been screened for BC specific mutations simultaneously with nuclear genome. Mutations observed at certain positions (204, 207 and 16293) of mtDNA that are obtained from nipple aspirate fluid have been an indicative for BC and mutations in mtDNA D-loop have been proposed as prognostic marker for BC independently. Even though there is a wide variation in the observed frequencies, region of mtDNA characterized and the nature of mutations associated with breast tumor tissue, it is obvious that mutations in mtDNA serve as informative

biomarkers for BC detection. Despite Mizoram, being a state with high incidence of BC and lifestyle of high tobacco consumption and unique dietary habits, data on the epidemiological and genetic risk factors is limited. To date, there have been no reports that focus on the genetics or environmental risk factors towards BC for the Mizo population. The present study on Characterization of clinically significant mutations associated with Breast Cancer in Mizo population has been proposed with the aim of evaluating the demographic risk factors and assess the prevalence of mutation in susceptibility genes associated with BC in Mizo population, Mizoram. The findings of this study may help us to better understand the causal risk factors and the prevalence of certain gene mutations that might further help in prevention, early diagnosis, and provide best treatment option to the patients.

**Objectives**

The following objectives are set forth to carry out the proposed work:

- Study of mutations in candidate genes and their association with BC in Mizo Population.
- To determine the potential confounding factors associated with BC and their pathogenicity analysis using in- silico methods.

## Materials and Methods

426 Breast Cancer (BC) patients from all over Mizoram and those registered in Mizoram State Cancer Institute (MSCI) within the year 2013 – 2017 and 810 healthy individuals were included in this study. 2 ml of blood was drawn from the participants (patients and healthy controls) by a trained technician and was stored in EDTA vials and kept in -20°C for further processing. Subject inclusion criteria are confirmed BC cases, with and without a history of BC in their family members, and BC patient of all ages from all over Mizoram, both registered as well as not registered in MSCI and belonging to Mizo tribe. Epidemiologic studies to determine the potential confounding factors associated with BC risk was performed by considering several factors such as reproductive history, environmental factors such as lifestyle and dietary habits, familial history of BC, other cancer and inheritable diseases. Association of histological tumor grade with other clinical and epidemiological features were statistical analysis using chi square test. To screen the genetic predisposition in Mizo BC patients, mutation in nuclear and mitochondrial DNA were screened using several approached such as direct Sanger sequencing and next generation sequencing which include targeted resequencing, clinical exome sequencing and whole mitochondrial genome sequencing.

**Results**

From the epidemiological studies of Mizo BC patients, it was found that the inheritable diseases, having 1st and 2nd degree relatives with BC, high consumption of khaini (tobacco), smoked food and sa-um (fermented pork fats), lack of physical exercises and shorter duration of lactation plays a major role in the development of BC. Late age at menopause, lesser intake of water and vegetables was found to act as potential confounding factors. Consumption of fish was also found to have an independent association with BC risk. From the formaldehyde estimation, fish coming from Andhra Pradesh was observed to have the highest concentration (3.17 µg/g), followed by Silchar (1.52 µg/g), Burma (0.28 µg/g) and Mizoram (0.17 µg/g).

From the analysis of association of histological tumor grade with other clinical and epidemiological features, a significant association was observed between tumor grade with lymph node status and tumor characteristics such as ER, HER2/neu, ER/PR, ER/PR/HER2neu. Tumor grade was also found to have an association with having 1st degree relative with BC. Interestingly, no association was observed between tumor grade with other tumor characteristics, reproductive and environmental factors that were considered in the analysis.

From the screening of mutation from six genes such as *BRCA1*, *TP53, PTEN, CDH1, CHEK2* and *XRCC2*, eight polymorphisms in three exons (11, 13 and 15) of *BRCA1* gene was identified. Four are non- synonymous and the other four are synonymous. Among the synonymous polymorphisms, one variant, g.95900A>T: c.4772A>T: p.P1544P in exon 15 was found to be novel. Several

genetics clinics and research groups had reported the other seven remaining polymorphisms in databases. There was no significant amino acid change to be noted since all the polymorphisms represent silent substitutions. In case of *TP53, PTEN, CDH1, CHEK2* and *XRCC2* gene, no genetic alterations was observed in the exons of interest.

From targeted re-sequencing of *BRCA1* and *BRCA2* whole gene analysis, two variants (c.T5089C:p.C1697R and c.C376T:p.Q126X) in *BRCA1* gene from sample 1F and 5F, respectively were found to be pathogenic. These two observed variants have been reported in databases for other populations and c.T5089C:p.C1697R is predicted to have a high score of pathogenicity according to ALIGN- GVGD grading. Among the studied samples, pathogenic variants were not observed in *BRCA2* gene.

From clinical exome sequencing, 16 variants consisting of 5- benign variants, 3- either benign or VUS (Variants of Uncertain Significance), 1- likely benign and 7 VUS were observed according to the ACMG (American College of Medical Genetics) variants classification guidelines. Also, from the observed variants, 3 variants in *BLM* (p.Thr162Ile), *CDKN2A* (intronic) and *FANCC* (p.Asp197Gly) were found to be novel. The variants found through the clinical exome sequencing were analyzed using the software prediction tools such as SIFT, Polyphen2, Provean, LRT, FATHMM, METASVM, METALR and CADD fetched us 12 deleterious variants. Majority of the variants are missense while only two variants are intronic-ss-acr and frameshift-ins.

From whole mitochondrial genome sequencing, a total of 119 non-synonymous variants were found in the samples studied (cases and controls).

Among them, 51 variants were observed in the protein coding region of mtDNA, in which 27 were specific to BC cases and 24 were found in common among the cases and controls. One frameshift variants (12014C>CA; c.1255_1255delinsCA) located in *ND4* gene was found to be novel. Four variants (*ATP6* 8860A>G, *ND3* 10398A>G, *ND5* 13708G>A and *ND5* 14110T>C) found in common for both cases and controls have been reported in database for BC. According to SIFT and Polyphen- 2, 22 variants were found to be affecting the normal protein function. Based on SOSUI software, 98% of non- synonymous variants fall in the hydrophobic region. The potential impact of a greater number of non-synonymous substitutions in the protein coding region represented decrease in protein stability, whereas, only two variant in ND2 and ND6 with ΔΔG of 0.144 and 0.025, respectively, were observed with increase protein stability according to Pmut.

68 variants in the non-coding displacement loop (d- loop) region of mitochondrial genome was observed, comprising of 26 BC specific variants and 42 common variants. The variant specific to BC cases, 16126T>C and other eight common variants such as 73A>G, 152T>C, 263A>G, 16189T>C, 16223C>T, 16290C>T, 16304T>C, 16362T>C has been reported for BC.

The ratio of non-synonymous to synonymous variants for mitochondrial respiratory complexes was analyzed and it was observed that the complexes vary in ratio. Complex V with 73% has the highest ratio in comparison to complex I, III and IV that were all lower than 40%. The dN/dS ratio of the variants in complex I, III, IV and V were 3.24, 0.894, 1.71 and 0.164, respectively. Since complex I and III were found to exceed 1, it indicates that there is a positive selection for non-synonymous variants in the two complexes in BC samples.

From the correlation analysis of 66 common variants present in mitochondrial genome, the frequencies of four variants were found to be significantly different between the cases and controls. Three of them were located in the coding region (ATP6 8701A>G, ND3 10398A>G and ND4L 10609T>C) and one in the d-loop region (16188 C>-T). These variants obtain a high odds ratio (> 1), which signifies that they are high risk factors for BC occurrence.

**Summary and Conclusion**

Breast cancer is the leading cancer site and major cause of cancer death in women and has become the global burden. This complex multifactorial disease regulated by genetic, hormonal and epigenetic factors such as diet and lifestyle habits has been an area of research interest. Decades of scientific research have lead us to the understanding of a fraction of the causal factors such as mutations in susceptibility nuclear as well as mitochondrial genes and also environmental factors which differs in different ethnic groups. Despite the groundbreaking advances being made in term of the discovery of more genes involved using high- throughput sequencing technologies and improve strategies for early detection and treatment options, there is a consistent increase in the prevalence of the disease.

The present study on Characterization of clinically significant mutations associated with Breast Cancer in Mizo population has been carried out with the aim of evaluating the demographic risk factors and assess the prevalence of mutation in susceptibility genes associated with breast cancer in Mizo population, Mizoram.

The findings of the present work are summarized as follows:

- In Mizoram, BC cases is observed from age as early as 20 and as old as 91 years. The mean average age was observed to be 49 years.
- Environmental factors such as inheritable diseases, having 1st and 2nd degree relative with BC, high consumption of khaini (tobacco), smoked food, sa-um (fermented pork fats), lack of physical exercises and shorter lactation duration were found to be the significant risk factors for BC in Mizo population.

- Other factors including late age at menopause, lesser intake of water and vegetables were also found to act as a potential confounding factors among the Mizo BC patient.
- We have observed that with the increase in the BC tumor grade, there is more of positive lymph node invasion. Tumor grade was found to have a significant association with lymph node invasion ($p<0.021$), ER ($p< 0.004$) and Her2/neu ($p< 0.014$) independently, double subtype of ER/PR ($p< 0.007$) and also the triple subtype ER/PR/HER2 ($p< 0.025$).
- The study showed that having at least one or more first degree relative's BC have a significant association with tumor grade ($p<0.003$).
- Interestingly, other variables such as tumor characteristics, reproductive factors, dietary and lifestyle habits does not show any significant association with tumor grade.
- Eight polymorphisms such as p.L771L, p.S694S, p.P871L, p.E1038G, p.K1183R, p.S1436S, p.P1544P, p.D1546Y were observed in BRCA1 gene from the candidate gene analysis using Sanger sequencing.
- Among them, p.P1544P in exon 15 was found to be novel. There was no significant amino acid change to be noted since all the polymorphisms represent silent substitutions.
- From targeted re-sequencing of BRCA1 and BRCA2 gene, two pathogenic mutations such as p.C1697R and p.Q126X in BRCA1 gene were observed in two familial Mizo BC samples. Among the familial BC samples included in the targeted re- sequencing, no pathogenic variants were observed in BRCA2 gene.

- Using clinical exome sequencing, the analysis based on ACMG (American College of Medical Genetics) variants classification guidelines, 16 variants consisting of 5- benign variants, 3- either benign or VUS (Variants of Uncertain Significance), 1- likely benign and 7 VUS were observed.
- Among them, 3 variants in BLM (p.Thr162Ile), CDKN2A (intronic) and FANCC (p.Asp197Gly) were found to be novel.
- The second approach for the analysis of clinical exome sequence data was performed using the software prediction tools such as SIFT, Polyphen2, Provean, LRT, FATHMM, METASVM, METALR and CADD. Twelve (12) deleterious variants in the Mizo population were observed: p.Arg2060Cys, p.Asn289His, p.Glu880Lys, p.Gly62Ser, CDKN2A (Intronic-ss-acr), p.Arg87Trp, p.Lys1358AspfsTer2, p.Ala370Val, p.Gly283Glu, p.Asp498Tyr, p.Thr363Ile, p.Arg181His.
- From whole mitochondrial sequencing, sequence alterations were found to be the highest in d-loop region and least in ND6 region.
- In the protein coding region, 51 variants were observed, in which 27 were BC specific and 24 were found in common among the cases and controls.
- One frameshift variants (12014C>CA; c.1255_1255delinsCA) located in ND4 gene was found to be novel. Four variants (ATP6 8860A>G, ND3 10398A>G, ND5 13708G>A and ND5 14110T>C) found common in both cases and controls have been reported in database for BC.
- According to SIFT and Polyphen- 2 software analysis, 22 variants were found to be affecting the normal protein functions. Based on SOSUI software, 98% of non- synonymous variants falls in the hydrophobic region.

- The potential impact of a greater number of non-synonymous substitutions in the protein coding region represented decrease in protein stability, whereas, only two variant in ND2 and ND6 with ΔΔG of 0.144 and 0.025, respectively, were observed with increase protein stability according to Pmut.
- In the non-coding d-loop region of mitochondrial genome, 68 variants comprising of 26 BC specific variants and 42 common variants was observed.
- The variant specific to BC cases, 16126T>C and other eight common variants such as 73A>G, 152T>C, 263A>G, 16189T>C, 16223C>T, 16290C>T, 16304T>C, 16362T>C has been reported for BC.
- The dN/dS ratio of the variants in complex I and III shows a positive selection for non- synonymous variants indicating the driver role of these complexes in BC development.
- According to correlation analysis, ATP6 8701A>G, ND3 10398A>G, ND4L 10609T>C in coding region and 16188 C>-T in the d-loop region were found to be high risk factors for BC occurrence in Mizo population.
- Analysis of variant clustering on mitochondrial genome based on the lifestyle risk factors showed that most of the familial BC cases falls under one major cluster and is observed to have great number of mitochondrial gene mutations.
- The major cluster showed familial BC and “tuibur” as the main risk factor. Dietary habits such as, high consumption of smoked food, Sa-um and fish are the major risk factor for BC.
- In case of non-familial BC cases, the frequency of mutations is elevated in mitochondrial respiratory complex 1 (ND1 – ND6).

The present study is the first scientific evaluation of the demographic, epidemiological and environmental risk factors towards BC progression and assessment of BC panel gene mutations in the population of Mizoram. The outcome of this study on several epidemiological risk factors can serve as baseline data to provide a more confined risk factors in different age groups of the population. Furthermore, mutation studies provide us with the knowledge on the prevalence of several BC susceptibility nuclear as well as mitochondrial gene mutations. Thus, from this study we can conclude that even though several reproductive, dietary and lifestyle factors play an important role in the development of BC, the extent of the contribution of several genes mutations need further validation in order to fully understand their role in BC progression in Mizo population.