# WHOLE GENOME ANALYSIS AND ITS PUTATIVE CLINICAL IMPACT ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

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


**ANDREW VANLALLAWMA**

**MZU REGISTRATION NO.: 1506957**

**Ph.D. REGISTRATION NO.: MZU/Ph.D. /1726 of 29.05.2020**




**DEPARTMENT OF BIOTECHNOLOGY**

**SCHOOL OF LIFE SCIENCES**


**MARCH 2024**

# WHOLE GENOME ANALYSIS AND ITS PUTATIVE CLINICAL IMPACT ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

**BY**

**ANDREW VANLALLAWMA**

**Department of Biotechnology**

**Supervisor: Prof. N. SENTHIL KUMAR**

**Submitted**

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

**MIZORAM UNIVERSITY**

**(A Central University)**

**Department of Biotechnology**

**Tanhril, Aizawl - 796004**

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

## CERTIFICATE

This is to certify that the thesis entitled **"Whole Genome Analysis and its putative Clinical Impact Associated with Breast Cancer in Mizo Population"** submitted by **Andrew Vanlallawma**, Ph.D. Scholar for the award of the Degree of Doctor of Philosophy in Biotechnology is carried out under my supervision and incorporates the student bona-fide research and this has not been submitted for the award of any degree in this or any other university or institute of learning.

Place: Aizawl, Mizoram
Date:

(PROF. N. SENTHIL KUMAR)
Supervisor

**MIZORAM UNIVERSITY**

**Month: March** **Year: 2024**

## DECLARATION

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

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

(ANDREW VANLALLAWMA)
Research Scholar

(PROF. JOHN ZOTHANZAMA)
Head of Department

(PROF. N. SENTHIL KUMAR)
Supervisor

## ACKNOWLEDGEMENT

It is a pleasure and privilege to express my sincere gratitude to my supervisor, Prof. N. Senthil Kumar, Department of Biotechnology at Mizoram University for his unwavering support, invaluable guidance, scholarly expertise and continuous encouragement. Without his presence, this work would never have reached a completion.

In addition, I would like to extend my sincerest appreciation to Prof. John Zothanzama, the Head of the Department of Biotechnology, and all the other faculties and non-teaching staffs for their unwavering support, encouragement during my research work.

I am thankful to all the Volunteers who took part and made it possible to conduct the study. A special thanks to Dr. Jeremy L. Pautu, Regional Cancer Research Centre, Zemabawk; Dr. John Zohmingthanga, Dr. Lalawmpuii Pachuau, Civil Hospital, Aizawl and Dr. Saia Chenkual, Zoram Medical College, Mizoram.

I am immensely grateful to my lab mates, Ranjan Jyoti Sarma, Zothanzami, Lalfakzuali, P.C. Lalrohlua Pachuau, Lalremmawia and all the research scholars of the Department of Biotechnology, Mizoram University, who have not only provided moral guidance but have also extended a helping hand whenever needed during my research work.

I would also like to give special recognition to David K. Zorinsanga for his immense help during my research work. Additionally, I am thankful to the Data Collectors and Lab Technicians, Jonathan Lalramhluna, K. Lalrampanmawii, T. Lalhriatpuii , Baby Lalrintluangi and R. Lalengkimi, for collecting the samples, clinical information, documentation, and DNA Isolation. I am also thankful to the Department of Biotechnology, New Delhi (GenomeIndia and DBT-Advanced State Biotech Hub) for the financial support during the research.

I would also like to thank all my family members, especially Fr. Caleb Laldawngsanga for his love and support during my Ph.D. course.

Lastly, I would like to express my profound gratitude to the almighty God for

granting me good health, strength, and unwavering determination throughout my research work.

**Place: Aizawl** **(ANDREW VANLALLAWMA)**

## TABLE OF CONTENTS


| Description | | Pages |
|---|---|---|
| *Certificate* | | *i* |
| *Declaration* | | *ii* |
| *Acknowledgement* | | *iii-iv* |
| *Table of Contents* | | *v* |
| *List of Tables* | | *ix* |
| *List of Figures* | | *x* |
| *List of Abbreviation* | | *xi-xiii* |
| **CHAPTER 1** | **INTRODUCTION AND REVIEW OF LITERATURE** | **1-9** |
| **CHAPTER 2** | **AIMS AND OBJECTIVES** | **10** |
| **CHAPTER 3** | **MATERIALS AND METHODS** | **11-21** |
| 3.1 | Sample Collection | 11-12 |
| 3.2 | Adverse Drug Reaction Information | 12 |
| 3.3 | Univariate Binary Logistic Regression | 12 |
| 3.4 | DNA isolation and Visualization | 13-14 |
| 3.5 | Whole Exome Sequencing and Data Analysis and Four Genes Targeted Panel | 14-16 |
| 3.6 | Whole Genome Sequencing and Analysis | 17 |
| 3.7 | Genes for identifying pharmacogenomic variants | 17-19 |
| 3.8 | Identification of pharmacogenomic variants | 19 |

| | | |
|---|---|---|
| 3.9 | Pathway analysis using DAVID | 20 |
| 3.10 | Identification of Druggable Genes | 20 |
| 3.11 | Inter-population differences in the allele frequency of important pharmacogenomic variants | 20-21 |
| 3.12 | Establishing common polymorphic variants among TNBC patients | 21 |
| 3.13 | Interpretation of population diversity using PCA plot | 21 |
| **CHAPTER 4** | **RESULTS** | **22-45** |
| 4.1 | Sample details | 22-23 |
| 4.2 | Representation of DNA Agarose Gel Electrophoresis picture | 23 |
| 4.3 | List of ADRs observed in TNBC patients | 24 |
| 4.4 | ADRs observed in TNBC patients with their severity | 25-26 |
| 4.5 | Total exonic variants observed after filtering | 27-30 |
| 4.6 | Distribution of variants in PharmGKB database, ClinVar and dbSNP | 31 |
| 4.9 | Variants that are present in PharmGKB Database | 32 |
| 4.10 | Variants present in ClinVar as pathogenic | 33 |
| 4.11 | Pathway analysis using DAVID (Kegg pathway) | 34 |
| 4.12 | Pathway analysis using DAVID (Reactome) | 35 |
| 4.13 | BRCA1: p.C1650R | 36 |
| 4.14 | DPYD (rs1801160),MTHFR (rs1801133), SLCO1B1 (rs4149056) and CYP2B6 (rs8192709) | 37 |
| 4.15 | Variants present in ClinVar but has no to little effect | 38-39 |
| 4.16 | Druggable categories and Drug Target | 40 |

| | | |
|---|---|---|
| 4.17 | Druggable targets and its Drugs | 41 |
| 4.18 | Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1 | 42 |
| 4.19 | Inter population differences in the allele frequency | 43 |
| 4.20 | Cataloguing of common polymorphism | 44 |
| 4.21 | Principal Component Analysis | 45 |
| **CHAPTER 5** | **DISCUSSION AND CONCLUSION** | **46-61** |
| 5.1 | Statistical analysis on ADRs with harboured mutation and clinical records | 46-48 |
| 5.2 | List of total variants observed (after variant prioritization) from TNBC patients | 48-49 |
| 5.3 | Variants that were present in PharmGKB database | 50 |
| 5.3a | MTHFR | 50-51 |
| 5.3b | SLCO1B1 | 52 |
| 5.3c | DPYD | 52-53 |
| 5.3d | CPY2B6 | 53-54 |
| 5.4 | Variants present in ClinVar | 55 |
| 5.4a | BRCA1 | 55-56 |
| 5.4b | VDR | 56-57 |
| 5.5 | Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1 | 57 |
| 5.6 | Inter-population differences in the allele frequency of important pharmacogenomic variants | 57-58 |
| 5.7 | Pathway analysis for Druggable target | 58-59 |

| | | |
|---|---|---|
| 5.8 | Cataloguing of common polymorphic variants among Mizo population | 59-60 |
| 5.9 | PCA analysis | 60-61 |
| **CHAPTER 6** | **SUMMARY** | **62-66** |
| **APPENDICES** | Questionnaire for Adverse Drug Reaction for Breast Cancer patients | **67-73** |
| | Ethics Committee Approval from Civil Hospital Aizawl | **74** |
| | Ethics Committee Approval from Mizoram State Cancer Institute, Zemabawk | **75** |
| **BIBLIOGRAPHY** | | **76-94** |

## LIST OF TABLES


| | | |
|---|---|---|
| **Table No.** | **Description** | **Page No.** |
| 3.1 | List of Triple Negative Breast Cancer recruited for the study. | 22-23 |
| 3.7a | Very Important Pharmacogenes (VIPs) extracted from PharmGKB. | 18 |
| 3.7b | List of Metabolizing Genes. | 19 |
| 4.3 | List of ADRs observed in TNBC patients. | 24 |
| 4.4 | ADRs observed in TNBC patients with their severity. | 25-26 |
| 4.5 | List of total exonic variants observed after filtering. | 27-30 |
| 4.9 | Variants that are present in PharGKB Database with their effects. | 32 |
| 4.15 | Representation of variants that were present in ClinVar but has no to little effects in disease conditions. | 38-39 |
| 4.16 | Drug-gene interactions observed from Drug Gene interaction database | 41 |

## LIST OF FIGURES


| Figure No. | Description | Page No. |
|---|---|---|
| 3.5 | WES analysis workflow till annotation. | 16 |
| 4.2 | Triple Negative Breast Cancer Genomic DNA gel picture of representative samples. | 23 |
| 4.8 | Distribution of variants observed after filtering. | 31 |
| 4.10 | Variant that are present in ClinVar with Clinical Significance as Pathogenic and Drug Response. | 33 |
| 4.11 | Pathway analysis using DAVID (KEGG pathway). | 34 |
| 4.12 | Pathway analysis using DAVID (Reactome). | 35 |
| 4.13 | Representation of one variant (*BRCA1:* p.C1650R) reported to be pathogenic in ClinVar. | 36 |
| 4.16 | List of Druggable Categories and Druggable target obtained by Drug Interaction function using Drug Gene Interaction database. | 40 |
| 4.18a | Targeted Panel Gene Sequencing of *BRCA1*, *BRCA2*, *CHEK2* and *PALB2* identifies 39 variants. | 42 |
| 4.18b | Variant location identified by Targeted Panel Gene Sequencing of *BRCA1*, *BRCA2*, *CHEK2* and *PALB2*. | 42 |
| 4.19 | Inter population differences in the allele frequency of important pharmacogenomic variants. | 43 |
| 4.20 | Cataloguing of common polymorphism with no clinical significance. | 44 |
| 4.21 | PCA of Mizo population with 1000 genomes. | 45 |

## LIST OF ABBREVIATION

| | | |
|---|---|---|
| BC | : | Breast Cancer |
| ASIR | : | Age Standardized Incidence Rate |
| ASMR | : | Age Standardized Mortality Rate |
| DCIS | : | Ductal Carcinoma In-Situ |
| IDC | : | Invasive Ductal Carcinoma |
| ILC | : | Invasive Lobular Carcinoma |
| TNBC | : | Triple Negative Breast Cancer |
| ER | : | Estrogen Receptor |
| PR | : | Progesterone Receptor |
| HER2 | : | Human Epidermal Growth Factor Receptor 2 |
| BLT | : | Basal Like Tmours |
| DNA | : | Deoxyribonucleic acid |
| CKs | : | Cytokeratins |
| CGH | : | Comparative Genomic Hybridization |
| BRCA1 | : | Breast Cancer Gene 1 |
| EGFR | : | Epidermal Growth Factor Receptor |
| DFS | : | Disease Free Survival |
| OS | : | Overall Survival |
| TCGA | : | The Cancer Genome Atlas |
| SNP | : | Single Nucleotide Polymorphism |
| GWAS | : | Genome Wide Association Study |
| mt-DNA | : | Mitochondrial DNA |
| ExAc | : | Exome Aggregation Consortium |
| ESP | : | Exome Sequencing Project |
| IGV | : | Indian Genome Variation |
| PBCR | : | Population Based Cancer Registry |
| NGS | : | Next Generation Sequencing |
| MSCI | : | Mizoram State Cancer Institute |

| | | |
|---|---|---|
| MBR | : | Modified Bloom Richardson |
| ADR | : | Adversed Drug Reaction |
| ETDA | : | Ethylenediamine tetra-acetic acid |
| CPG | : | Clinical Practice Guidelines |
| NCCN | : | National Comprehensive Cancer Network |
| TNM | : | Tumour – Node - Metastasis |
| WES | : | Whole Exome Sequencing |
| SPSS | : | Statistical Package for Social Sciences |
| RT | : | Room Temperature |
| TAE | : | Tris acetate- EDTA |
| UV | : | Ultra Violet |
| WEAP | : | Whole Exome Analysis Pipeline |
| GATK | : | Genome Analysis Tool Kit |
| BWA-MEM | : | Maximal Exact Match |
| PCR | : | Polymerase Chain Reaction |
| VCF | : | Variant Calling Format |
| ANNOVAR | : | Annotate Variation |
| SIFT | : | Sorting Intolerant from Tolerant |
| DAVID | : | Database for Annotation, Visualization and Integrated Discovery |
| KEGG | : | Kyoto Encyclopedia of Genes and Genomes |
| PCA | : | Principal Component Analysis |
| VIP | : | Very Important Pharmacogenes |
| MTHFR | : | Methylenetetrahydrofolate reductase |
| DPYD | : | Dihydropyrimidine dehydrogenase |
| SCLO1B1 | : | Solute Carrier organic anion transporter family member 1B1 |
| CYP2B6 | : | Cytochrome P450 Family 2 Subfamily B Member 6 |
| HR | : | Homologous Recombination |
| NHEJ | : | Non-Homologous End Joining |
| OC | : | Ovarian Cancer |

| | | |
|---|---|---|
| GC | | Gastric Cancer |
| dUMP | | deoxyuridine monophosphate |
| dTMP | | deoxythymidine monophosphate |
| SLC | | Solute Carrier |
| DPD | | Dihydropyrimidine dehydrogenase |
| 5-Fu | | 5- Fluorouracil |
| CPIC | | Clinical Pharmacogenetics Implementation Consortium |
| SGE | | Saturation Genome Editing |
| VDR | | Vitamin D Receptor |
| mRNA | | messenger Ribonucleic Acid |
| ADME | | Absorption Distribution Metabolism and Excretion |

# Chapter 1

## Introduction and Review of Literature

The human body has approximately 36 and 28 trillion cells in males and females, respectively (Hatton *et al.,* 2023). These cells grow and divide to give new cells as per the body's requirement and when the cells become aged or functionally impaired, they die and new cells replace them. Occasionally, this orderly progression may malfunction, leading to an inappropriate proliferation of abnormal cells or evading programmed cell death. These cells might develop into tumours, which are abnormal masses of tissue. These tumours can be either cancerous or non-cancerous (benign). Cancerous cells are characterised by their ability to invade and spread to nearby tissues or at a distant place within the body to form new cancer cells, a process called as "metastasis" (Hanahan and Weinberg, 2000). In other words, cancerous cells are called as malignant cells which usually form solid tumours. However, there are other types of malignancy that do not form solid tumours such as cancer of the blood like Leukemia.

Cancer represents a significant global health challenge, with millions of new cases diagnosed each year (Bray *et al.,* 2018). Cancer is not a single entity but encompasses a wide range of diseases and can affect virtually any tissue or organ, giving rise to a diverse array of malignancies, each with distinct etiologies, pathologies, molecular characteristics and treatment strategies. According to Globocan (2020), the global incidence and mortality of all cancer was registered to be 19,292,789 and 9,958,133, respectively where 49.26 % are Asians. Breast Cancer (BC) has the highest Age-Standardized Incidence Rate (ASIR) that is 47.8 per 100,000 and second highest Age-Standardize Mortality Rate (ASMR) accounting to 13.6 per 100,000. (Sung *et al.,* 2020). In India, BC ranks first both in ASIR and ASMR with 28.5 and 13.3 per 100,0000, respectively. Among female, the recorded new cases of cancer indicates that BC was 1,78,361 accounting to 26.3% (Sung *et al.,* 2021).

The breast has mainly glands, ducts, and fatty tissue that resides on top of the upper ribs and chest pectoral muscles. BC can start from any parts of the breast and according they are named as such. Lobules are the glands that produce breast milk, lobular cancers are those when the lobular cells turn malignant. Similarly, Ducts, are small vessels that connects the lobules to the nipple and carries breast milk. and ductal cancers are the most common among breast cancers. The areola is a thicker skin that composed the nipple, where the ducts converge and forms a larger vessel that carries the breast milk and this when turns malignant are called as Paget disease of the breast and are lesser-known form of BC. The fat and connective tissue (stroma) envelop the ducts and lobules, providing structural support and when these stromal cells become cancerous, they are called as Phyllodes tumor which is also a rare type of breast cancer, Blood vessels and lymph vessels are also present in both the breast, with angiosarcoma occurring less frequently in breast cancer that originates in the lining of these vessels (Henry *et al.,* 2020).

The majority of BC are classified as carcinomas, a type of tumour originating from epithelial cells lining various organs and tissues in the body. In the case of BC, the predominant type is adenocarcinoma, which initiates in the cells of either the milk ducts or the lobules (milk-producing glands) within the breast. The type of specific cells within the breast that turns malignant dictates the type of BC. The type of BC can also be an indication of whether it had metastasised or resides locally. (American Cancer Society)

According to the site, BC can be classified accordingly to Ductal Carcinoma In-Situ (DCIS) and Invasive/Infiltrating BC. Such classification is based upon whether the malignant cells are residing locally or it had metastasised. DCIS is a type of BC where malignant cells are enclosed/resides within the duct and does not spread to the surrounding tissues. DCIS are also referred as non-invasive or pre-invasive BC, however, they can turn into invasive which can later metastasised. While, Infiltrating BC are those that metastasised into the surrounding tissues. They can be further classified in Ductal and Lobular. Infiltrating Ductal Carcinoma (IDC) as the name suggests, the lining of the milk-duct turns cancerous and these malignant cells invades the neighbouring tissues, while in Infiltrating Lobular Carcinoma (ILC),

malignant cells originate from the milk producing glands called the lobule. The incidence of IDC is more common, 8 out of 10 BC than ILC which is 1 in 5 of BC (Sharma *et al.,* 2010).

Certain invasive BC exhibit distinctive characteristics or progress through differently, impacting their treatment approach and prognosis. While these types of cancers are less prevalent, they can pose a more significant threat compared to other forms of breast cancer. Such type of invasive BC is the Triple Negative Breast Cancer (TNBC), which is aggressive, lacks Estrogen Receptors (ER-), Progesterone Receptors (PR-), and Human Epidermal Growth Factor Receptor 2 (HER2-negative).

Although there are many sub types of BC, some of which are common like DCIS and IDC, some have special features like TNBC and Inflammatory Breast Cancer while some are less common such as Paget disease of the Breast Cancer and Phyllodes tumour. However, there are many studies that categorises the different types of BC using similar gene expression by micro array technique (Perou *et al.,* 2000) and was validated by different research groups (Sotiriou *et al.,* 2003). These findings were also expanded by different independent research groups and BC was classified into five distinct molecular groups. These classifications were then broadly classified into two broad groups after cluster analysis as: Estrogen receptor positive (ER-positive) and Estrogen receptor negative (ER-negative) groups with their own subtypes (Sørlie *et al.,* 2001; 2003; Zhao *et al.,* 2004).

ER-positive subtypes can be classified according to which expresses ER and genes from the luminal epithelial cells. While the subclassification of luminal tumours remains a topic of debate, there appear to be at least two categories of ER-positive breast cancers. These are typically referred to as luminal A and B tumours, determined by the extent of characteristic gene expression and the presence of other genes associated with either the proliferation cluster with or without HER-2 expression. While ER-negative can be accordingly grouped as those that are ER-negative and are classified into three subtypes as HER-2-positive, Basal-like Tumours and Breast–like Tumours (Calza *et al.,* 2006; Hu *et al.,* 2006).

Basal-Like Tumours (BLT) showed absence of ER and HER-2 expression and have a distinct clinical phenotype as this classification is based on gene expression that have a prognostic value. These BLT showed wide range of gene expression such as Luminal Cytokeratins (CKs), *MUC1, BCL2, GATA3,* calponin, p63, integrin, CD10, P-cadherin, etc. (Rakha *et al.,* 2008) and this confers a greater challenge in predicting treatment strategies where all these classifications were based on gene expression to provide an insight to have a better understanding of what drugs to be administered.

Parallel to gene expression studies, microarray-based comparative genomic hybridization (array-CGH) techniques were employed to study the genomic alterations and were found to have a greater genomic instability than other BC types with low levels of deletions and amplification (Bergamaschi *et al.,* 2006; Chin *et al.,* 2006; Vincent-Salomon *et al.,* 2007). As such was the case, BLT was predicted to have more of Deoxyribonucleic Acid (DNA) double stand breaks in the DNA repair mechanism such as Breast Cancer Gene 1 (*BRCA1)* (Turner *et al.,* 2006) and Epidermal Growth Factor Receptor (*EGFR)* (Reis-Filho *et al.,* 2006).

Although in the previous classification, which classified only with ER-positive or ER- negative, BLT were apart of ER- negative group. One of the early references of TNBC is often attributed to the work of Carey *et al.,* (2006) which was previously referred to as basal – like subtype (Perou *et al.,* 2000). However, as per Lehmann *et al.,* 2011, TNBC was classified into six subtypes: immunomodulatory, mesenchymal stem-like, mesenchymal, luminal androgen receptor, basal-like 1 and basal-like 2, which they later re-classify to four subtypes mainly to BL1, BL2, M and LAR (Lehmann *et al.,* 2016). Other study groups also classified into four types such as LAR, M, BLIS, and BLIA (Burstein *et al.,* 2015), while in 2021, FUTURE trial schema classified as LAR, IM, MES and BLIS (Jiang *et al.,* 2021). This suggested that classifying TNBC using gene expression method is ever changing and evolving. This merits the need for identifying genomic mutations amongst this classification that could also be used as a diagnostic or prognostic, predict Disease Free Survival (DSF) and Overall Survival (OS).

Apart from the differential gene expression from other types of BC, mutational signatures or prone to TNBC was also observed in previous studies. Like many of cancers, *TP53* is also prone to mutation that regulates cell cycle and apoptosis (Stover *et al*., 2018; Shah *et al*., 2012) and *PIK3CA* gain-of-function mutation (Shah *et al*., 2012; Cancer Genome Atlas). It is well established that *BRCA1* germline mutation accounts for approximately 15% in TNBC patients (Sharma *et al*., 2014) which is then frequently associated with somatic mutation in the TP53 gene (Greenblatt *et al.,* 2001), this is probably attributed, at least partially, to the *BRCA1* protein's capacity to engage with and control the transactivation of target genes regulated by p53 (Zhang *et al.,* 1998). *KMTC* is most prone to Thai population upon comparison with The Cancer Genome Atlas (TCGA), Molecular Taxonomy of Breast Cancer International Consortium (METABRIC) where samples were collected from United Kingdom and Canada and the French cohort (Niyomnaitham *et al*., 2019). Other study groups also reported TNBC predisposing genes in *PALB2*, *FANCM* (Cybulski *et al.,*2015; Neidhardt *et al*., 2017). *RAD51D*, and *ATM* were also considered to be highly mutated among 1104 TNBC patients (Sun *et al*., 2017). Similarly, from 1824 TNBC patients (Couch *et al.,* 2014) identified *BARD1, BRIP1, RAD51C, RAD51D, RAD50,* and *XRCC2* as BC susceptible genes. It is evident that from the mentioned studies, even though with little references, different population has distinct mutational signature in TNBC patients. This indicates the genome complexity within TNBC which is associated with an aggressive tumour and that fact that it lacks molecular targets makes it more difficult to obtain better treatment that it relies on chemotherapeutic treatment modalities (Shi *et al.,* 2018).

The success or failure of cancer treatment outcomes hinges on various factors, including co-morbidities, age, gender, cancer type, stress levels, and the specific drugs administered. These factors, coupled with an individual's genomic profile, collectively determine the drug's effectiveness (Shargel *et al*., 2005). Precision medicine emerges as a solution to address adverse effects in critical conditions by tailoring treatments to the unique genetic makeup of individuals. This approach aims to identify the most suitable drugs for each person, optimizing efficacy while minimizing toxic effects (Gonzalez-Angulo *et al*., 2010).

Variations in the genome that directly impact the way drugs are processed (pharmacokinetics) and their effects on the body (pharmacodynamics) are referred to as pharmacogenetic variations. These variations play a pivotal role in determining clinical outcomes, and the markers associated with these outcomes are known as biomarkers. Specifically, biomarkers linked to a particular clinical outcome are termed prognostic biomarkers, and those that can guide treatment strategies are referred to as predictive biomarkers (Sacco and Grech, 2015).

In the context of breast cancer, well-established prognostic biomarkers that inform treatment strategies include estrogen receptor (ER), progesterone receptor (PR), and human epidermal growth factor receptor-2 (HER-2) (Baldacchino *et al*., 2014). For instance, patients who are ER-positive can benefit from selective estrogen receptor modulators (SERMs) such as tamoxifen or aromatase inhibitors. On the other hand, HER2-positive patients can be treated with trastuzumab to inhibit tumour growth (Mandrekar and Sargent, 2009). These biomarker-guided approaches contribute to more personalized and effective treatment strategies for breast cancer patients.

There are different genes that plays a role in the outcome of the treatment administered. These genes when there are changes within them usually Single Nucleotide Polymorphisms (SNPs) there are referred as pharmaco-gene variants. One of the famous gene that is a pharmacogenic marker of tamoxifen and its variants is the *CYP2D6* gene and depending on the *CYP2D6* variants, patients have been categorized as extensive metabolizers (EM), intermediate (IM), or poor metabolizers (PMs) as multiple allelic variants *3, *4, *5, shows negligible enzyme activity while *9,*10,*17 shows decrease in the enzyme activity (Jin *et al*., 2005; Bradford, 2002; Goetz *et al*., 2007).

*CYP19A1* is a pharmacogenic marker of aromatase inhibitors that are involved in the efficacy and its toxic effects to the patients (Colomer *et al*., 2008) Similarly, different SNPs are identified in *CYP2D6* (Ma *et al*., 2005) and *TCL1A* (Ingle, 2010) using Genome-wide association studies (GWAS) but are also

population specific suggesting the need to identify a population specific pharmacogenetic variants.

Paleontological and genetic evidence traced back the human origin from Africa within the past three hundred thousand years (ky) (Hublin *et al*., 2017) and spread across the globe within the last 100 ky (Campbell et al., 2014). Evolutionarily, the Indian subcontinent has been a corridor for different migratory waves arising from Africa, through land as well as coastline routes (Reich *et al*., 2009; Majumder and Basu, 2014). However, Mizoram being one of the states in India seems not to be the migratory route and that the people of different parts of India looks phenotypically different (Majumder and Basu, 2014). Mizo population has no such scientific genomic study on its genealogy. Population based genomic study can reveal the genetic make-up, its migration pattern and inheritance of certain polymorphisms that could have implications to certain diseases as well as a reference of the Origin of Mizo's for the state's archive collection.

The Indian subcontinent is the second largest populated country in the world with huge genetic diversity. This diversity can be classified into 4 major groups on the basis of ethno-racially as Caucasoids, Australoids, Mongoloids and Negritos. Linguistically, Indians can be grouped into four categories as Indo-European (north), Dravidian (south), Tibeto-Burman (north east) and Austro-Asiatic (central and east) (Indian Genome Variation Consortium, 2005). This diversity can also be attributed to India as being a pathway for human migration pattern that originates from Africa to India and the rest of the world through land and coastal routes (Reich *et al.*, 2009; Majumder and Basu., 2014). The genetic variations that are observed in Indians are also moderately mirrored in the mitochondrial DNA (mtDNA), Y-Chromosome and certain genes or markers (Bamshad *et al.,* 2001; Kivisild *et al.,* 2003; Das *et al.,* 1996; Majumder *et al.,* 1999; Thanseem *et al.,* 2006).

However, these findings majorly are cumulative of small datasets from different scientific work and among them the largest scientific study was from the Indian Genome Variation Consortium where they have identified single nucleotide polymorphisms from 900 genes out of 1800 Indian individuals (Indian Genome

Variation Consortium, 2005). Recently, IndiGenome has performed whole genome sequencing of over 1000 Indian samples and has catalogued variants within the population and found 37,249,254 (66.63%) variants which are common with gnomAD database and 21,485,966 (38.43%) variants with the 1000 Genomes project. There were 20,853,355 (37.30%) variants common in all the three databases whereas 18,016,257 (32.23%) variants were unique to only IndiGenomes database (Jain *et al.,* 2020).

In the post genomic era, with the development of high throughput sequencing platforms and its substantial decrease in the cost has led to an extensive flare-up of sequencing projects at a global level or at population specific (Genome of the Netherlands Consortium, 2014; 1000 Genomes Project Consortium, 2015; Gurdasani *et al.,* 2015; Mallick *et al.,* 2016; Nagasaki *et al.,* 2015; Scott *et al.,* 2016; Fakhro *et al.,* 2016). Certain population specific datasets including ExAC (Lek *et al.,* 2016), gnomAD (Karczewski *et al.,* 2019), ESP6500 (Exome Variant Server: (https://evs.gs.washington.edu/EVS/) and 1000 genome project (1000 Genomes Project Consortium, 2015) have included individuals from India, but this does not justify the true nature of diverse Indian population (Sengupta *et al.,* 2016). Indian Genome Variation (IGV) consortium identified SNPs from 1800 Indians revealing unique founder mutations leading to better understanding of genomic variants and its disease phenotype (Indian Genome Variation Consortium, 2008).

Mizoram has the highest cancer cases in India and are being dubbed as "Cancer Capita of India" by Times of India in December 2016 (https://www.indiatimes.com). The incidence of cancer in Mizo females as per 100,000 is 747 and 380 for Mizoram State and Aizawl, respectively within the year 2012 – 2016 with a mortality rate of 76.4% and 89.8%, respectively and a cumulative risk of 1.5 and 1.5, respectively (Prashant *et al.,* 2020). These Population Based Cancer Registries (PBCR) report poses an important public health problem and is imperative that more scientific study has to be conducted with the population to abate this huge burden.

With the advancement of Next Generation Sequencing (NGS), huge datasets are generated with the intention of creating a population specific variant catalogue that can play a crucial role in clinical applications. This huge data generated needs a better-quality control for it to have a clinical use, therefore, it is imperative that each ethnic background has their own systematic classification of variants that are of basal polymorphisms and those variants that play an important role in medical setups. This study aims to understand the pharmacogenomically important variants among TNBC patients in Mizoram and also to create a reference of common polymorphism in healthy individuals that can later be use as a filtering criterion for analysing huge NGS data to fish out the disease causal variant or important pharmaco-variants.

It is clearly evident from the cross-sectional study of PBCR from 2003 to 2020, that Mizoram has one of the highest incidences of Cancer within the country. This can be attributed to lack of early diagnostics, no mutational screening panel, infrastructure, less scientific exploration, skilled man power, transportation problem etc. (Zomawia *et al.,* 2023). We also perform a join-point analysis of the ASIR and ASMR of all BC patients that were reported in PBCR from 2003 to 2022 and found that the Annual Percent Change of both ASIR and ASMR were 3.60 and 7.11 respectively, suggesting the need for scientific research within the State. This study aims to identify the genomic background of TNBC patients in Mizoram and its uses in pharmacogenomics. As genomic sequencing generates huge data, we also aim to identify the common polymorphisms and catalogue it for in-house data filtering step. We also will study the genomic differences that occurs between other populations and identify the closest populations so as to make an informed knowledge on how to approach genomic difficulties with respect to pharmacogenomics and personalized medicine.

## Chapter 2

## Aims and Objectives

1. Identification of genetic variants associated with pharmaco-genes in Triple Negative Breast Cancer.

2. Interpretation of population diversity using common polymorphic variants of Mizo tribal population.

# Chapter 3

# Materials and Methods

### *3.1. Sample collection*

In order to conduct this current study, Institutional Ethics Committee approval was obtained from Mizoram State Cancer Institute (MSCI)-No.D.12016/2/2013-MSCI/IEC (Appendix I). A total of 28 TNBC patients who are admitted in MSCI were recruited for the study after obtaining a signed consent form. The TNBC patients recruited for the study has a median age of 52 years with the oldest being 91 years and youngest being 24 years, providing a wider range for studying if age has an effect in the efficacy of chemotherapy drugs administered to them (Table. 3.1). Their clinical records such as histopathology, tumour grade, Modified Bloom-Richardson (MBR) grade, axillary lymph node involvement, chemotherapeutic treatment administered and information regarding Advanced Drug Reaction (ADR) were also collected using a well-structured questionnaire. Peripheral blood of 5 ml was drawn from each TNBC patients to study the genomic alterations and its implication in manifestation of different ADR's. The blood sample was collected in Ethylenediamine tetra-acetic acid (EDTA) coated vials and was stored in $-20^0$C until it was ready for DNA isolation.

Similarly, 20 individuals who are self-proclaimed healthy were taken as healthy controls with the possibility of using their genomic data to study the diversity with other populations. These 20 individuals were carefully selected based on: genetically unrelated, different sub-tribes representing the different districts in-order to maintain sample diversity. These 20 individuals are from a sub-set of a project entitled "Genomics for Public Health in India (IndiGenomes)" initiated by CSIR-Institute of Genomics and Integrative Biology, New Delhi. This pan-India project aims to understand the human genome and diseases from Indian population. Institutional Human Ethics Committee of Civil Hospital Aizawl approval was obtained- No.B.12018/1/13-CH(A)/IEC/94 (Appendix II), and a signed inform

consent from the volunteers, peripheral blood of 5 ml was collected in EDTA vials and stored in -20$^0$C.

### *3.2. Adverse Drug Reaction Information*

This study aims to corelate between the genomic background of TNBC patients with the manifestations of different ADRs and their clinical records. To study the possible ADRs with different treatment regime, a well-structured questionnaire was formulated to follow the uniform Clinical Practice Guidelines (CPG). For this study we have studied different CPGs from different countries and opted to follow the National Comprehensive Cancer Network (NCCN). The structured questionnaire contains information such as personal details, date of diagnosis, histopathology, tumour location and grade, MBR grade, auxiliary lymph node involvement, types of chemotherapy drugs administered and no of cycles, Tumour – Node - Metastasis (TNM) staging and different types of ADRs. The questionnaire was followed-up in 6 months interval to check their vital status and to record any complications they had endured.

### *3.3. Univariate Binary Logistic Regression*

The questionnaires content such as clinical and histopathological records like tumour grade, MBR grade, auxiliary lymph node involvement, TNM staging, the combinations of chemotherapy administered and the number of cycles can affect the drug's efficacy. Similarly, Whole Exome Sequencing (WES) analysis also reveals variants that are of pharmacogenomic importance, also there are variants that are also reported and classified as pathogenic in ClinVar database. In order to understand the relation of genomic variants, clinical and histopathological records and ADR, we performed test to identify whether our variables were normally distributed. Since, our data was not normally distributed, Spearman rank correlation co-efficient was used for significant level (p-value) $<0.05$ and confidence level of 95% were maintained using IBM Statistical Package for Social Sciences (SPSS) version 22.0.

### *3.4. DNA isolation and Visualization*

For Whole Exome Sequencing of TNBC patients, DNA isolation was performed using QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN) as per the manufacturer's protocol with little modifications. The stored blood of 5 ml was thawed and water-bath was set to $56^0$C. QIAGEN protease and buffers AW1 and AW2 were prepared. 20 µl of protease was pipetted out into a 1.5 ml microcentrifuge tube and 200 µl of whole blood was added to it along with 200 µl of AL buffer. The solution was pulse vortexed for 30 seconds and was incubated at $56^0$C in water-bath for 30 mins which was then inverted five times for every 5 mins. After incubation, the solution was pulse vortexed for 30 seconds and 200 µl of chilled ethanol was added to it, followed by pulse vortex for 30 seconds. The entire lysate was transferred into a QIAamp Mini-spin column placed in a 2 ml collection tube and was vortexed at 6000xg for 1min at room temperature (RT). The 2 ml collection tube was then discarded and a new collection tube was used. 500 µl of AW1 buffer was added to the spin column and was centrifuged at 6000xg for 1min at RT, after which the 2 ml collection tube was discarded and was replaced by a new 2 ml collection tube. 500 µl of AW2 buffer was added to the QIAamp Mini spin column and was centrifuged at 20,000xg for 3 min at RT. The flow through was discarded and the same QIAamp Mini spin column was put in the 2 ml collection tube which was centrifuged at 20,000xg for 1 min at room temperature. The 2 ml collection tube was discarded and a new 1.5 ml microcentrifuge tube was used and dried in Laminar Air Flow chamber for 15 mins. 200 µl of AE buffer was directly added to the membrane of the spin column and incubated for 10 mins and centrifuged at 6000xg for 1 min at RT, and the eluted genomic DNA was stored at $-20^0$C deep freezer.

The eluted genomic DNA was visualized by Agarose Gel Electrophoresis. In a 100 ml conical flask, 80 ml of Tris acetate- EDTA (TAE) buffer and 0.64 g of Agarose was added and boiled in the microwave oven. 4 µl of Ethidium Bromide was added when the solution cools down to a Luke warm solution and was poured on the gel electrophoresis casting tray and allowed it solidify for 45 mins at room temperature. 4 µl of genomic DNA and 0.5 µl of loading dye was mixed together and was loaded in the wells of the agarose and 3 µl of 100bp DNA ladder was used as

marker and was electrophoresed for 45 mins and visualized in UV Gel documentation system (Bio-RAD)

### *3.5. Whole Exome Sequencing and Data Analysis and Four Genes Targeted Panel*

About 17 TNBC samples were sequenced at the National Institute of Biomedical Genomics, Kalyani, India and for library preparation, Agilent SureSelectXT target enrichment System was used following the manufacturer's protocol and sequenced in Illumina Hi-Seq2500 at 100X depth.

WES data and Targeted Panel analysis was performed using an in-house developed automated pipeline called Whole Exome Analysis Pipeline (WEAP) (In-house pipeline) following the best practices outlined by Genome Analysis Tool Kit (GATK) guidelines. The pipeline WEAP utilizes various existing tools following GATK pipelines such as BWA aligner (Li and Durbin, 2009), Samtools (Li *et al*., 2009), Picard (http://broadinstitute.github.io/picard.), and ANNOVAR (Wang *et al*., 2010). WEAP has a unique feature that allows quality checking of the raw reads and simultaneously trim the low-quality reads and checks the quality parallelly using fastp tool. Trimmomatic was also used for removing low quality reads (Bogler et al., 2014). The WES analysis workflow till annotation is presented in Figure. 3.5.

*Reads Alignment and Removal of PCR Duplicate:* BWA-MEM was used to aligned the raw trimmed files which had already passed the quality checking, against the reference genome - GRCh38.p13 generating a SAM file which was converted to BAM file. The BAM file was then sorted using samtools utility according to their respective genomic coordinates. As NGS is based on fragmentation of DNA into multiple fragments and is then amplified by a process known as bridge Polymerase Chain Reaction (PCR), there is a possibility that same fragments are originating from the same fragment upon alignment. These are removed by using Picard tool as they can provide a false positive information

*Base Quality Score Recalibration (BQSR):* Similar to the quality checking for the raw fastq and the trimmed files, quality checking of the aligned and removed

duplicates BAM file is required and was performed using GATK Baserecalibrator utilizing a known variant dataset, "dbSNP138.vcf," to recalibrate base quality scores.

*Germline Variant Calling and Recalibration:* To identify nucleotides that are not matching with the reference genome (called as variants) GATK HaplotypeCaller was used. This tool is used to identify variants present in germline as our sample was whole blood. To accurately call germline variants GATK VariantRecalibrator was used to remove sequencing artifacts (Van der Auwera *et al*., 2020).

*Variant annotation:* The variants that passed quality check is in a Variant Call Faormat (VCF) format. This file contains few details of the variants such as chromosome location, the variant position, the reference allele, the alternate allele and some others. To have a better understanding of a variant, annotation by ANNOVAR was performed. ANNOVAR utilises a number of publicly available databases such as RefGen, avsnp15, exac03, esp3500, ljb26, gnomad211exome, ClinVar, and 1000g. Such databases have their own niche of information such as variant location, name of gene, function of gene, allele frequency, clinical impact, pathogenicity prediction etc.

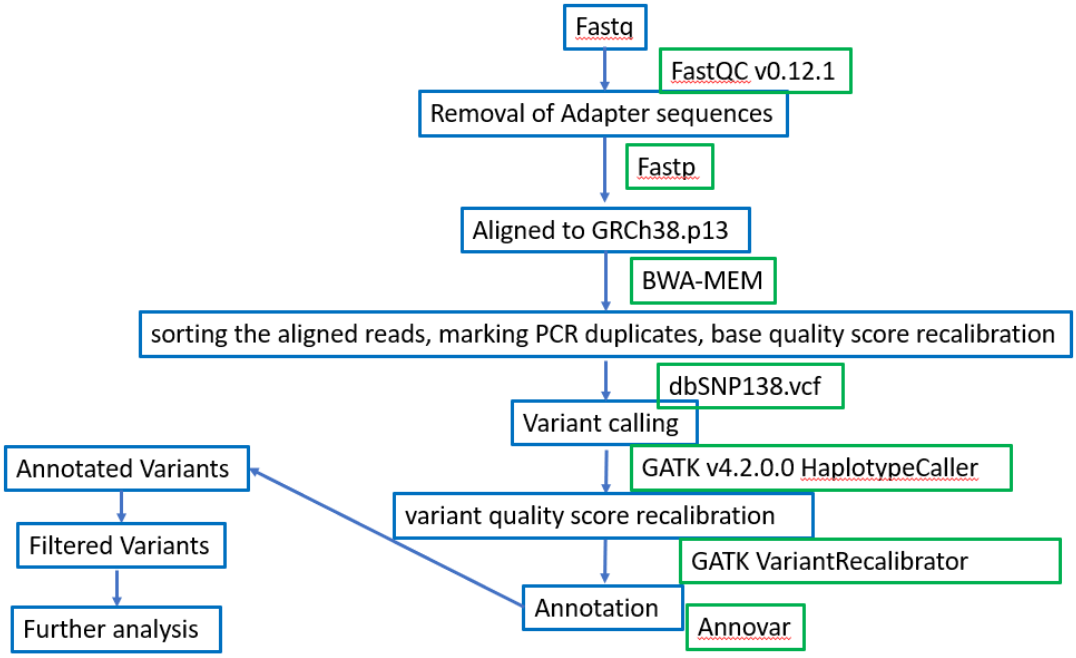



Figure. 3.5: WES analysis workflow till annotation.

### *3.6. Whole Genome Sequencing and Analysis*

Whole Genome Sequencing was performed at CSIR-Institute of Genomics and Integrative Biology, New Delhi as a collaborative effort. The libraries were prepared using TruSeq DNA PCR free library preparation kit as per manufacturer's instructions (Illumina Inc. San Diego, CA, USA, Cat. no. FC-121-9006DOC). Sequencing of the prepared libraries was performed on Illumina NovaSeq 6000 platform (Illumina Inc. San Diego, CA, USA). Alignment to reference genome, refining and default quality filtered variant calling was performed on the Illumina DRAGEN v3.4 Bio-IT platform (Illumina Inc. San Diego, CA, USA) usingGRCh38 as a human reference genome. Sentieon was used for joint variant calling that mimic the tools present in GATK and also calculates the allele frequency, allele count and allele number for the observed variants.

### *3.7. Genes for identifying pharmacogenomic variants*

WES annotated file contains approximately 30,000 variants per individual. This suggests that there are lots of variants that can be classified as variants of insignificance based on the impact towards contributing to disease manifestation, disease progression, susceptibility to, protective, risk factor and drug metabolism. Such variants can be classified as synonymous variants (variants that do not change the amino acid), variants present in intronic region, polymorphisms that present both in diseased condition and in normal with high allele frequency and many more. To avoid such all variants, filtering steps were applied. The first filtering step used was "The Very Important Pharmacogenes" which is listed out in the PharmGKB database (later referred to as PharmGKB gene list) (Table. 3.7a.). The list provides a well curated genes that confers a significant role in the metabolism or response to one or a combination of multiple drugs. The list contains 69 genes while analysis on this study was being conducted.

Table. 3.7a..: Very Important Pharmacogenes (VIPs) extracted from PharmGKB database.

| **PharmGKB VIP Genes: 68** | | | | |
|---|---|---|---|---|
| ABCB1 | BRCA1 | CYP3A4 | KCNJ11 | SLC22A1 |
| ABCG2 | CACNA1S | CYP3A4 | KIT | SLCO1B1 |
| ABL1 | CFTR | CYP3A5 | KRAS | SULT1A1 |
| ACE | COMT | CYP4F2 | MTHFR | TPMT |
| ADH1A | CYP1A2 | DPYD | MT-RNR1 | TYMS |
| ADH1B | CYP2A13 | DRD2 | NAT2 | UGT1A1 |
| ADH1C | CYP2A6 | EGFR | NQO1 | VDR |
| ADRB1 | CYP2B6 | ERBB2 | NR1I2 | VKORC1 |
| ADRB2 | CYP2C19 | F5 | NRAS | |
| AHR | CYP2C8 | G6PD | NUDT15 | |
| ALDH1A1 | CYP2C9 | GSTP1 | P2RY12 | |
| ALK | CYP2D6 | GSTT1 | PTGIS | |
| ALOX5 | CYP2D6 | HLA-B | RYR1 | |
| BCR | CYP2E1 | HMGCR | SCN5A | |
| BRAF | CYP2J2 | KCNH2 | SLC19A1 | |

In this study, majority of the TNBC patients received almost the same combination of chemotherapy drugs such as 5-Fluorocil, Epirubicin, Cyclophosphamide and Paclitaxel with Capecitabine, Doxorubicin, Docetaxel as a substitute in few of the patients. All the genes involved in transport and metabolizing these drugs were listed. Apart from this list of genes, genes that have pharmacogenomic importance that are listed in DrugBank was also utilized. DrugBank provides a list of genes that involves in the metabolism and transport of drugs. Finally, genes that were still absent in this list but have importance were obtained from literature survey which was later overlapped with DrugBank database. This list contains 88 genes (later referred to as Metabolizing gene list) (Table. 3.7b.).

Table. 3.7b.: List of Metabolizing genes

| Metabolizing genes: 88 | | | | | |
|---|---|---|---|---|---|
| ABCB | ATG7 | CYP2B6 | ERBB2 | MTHFD1 | SLC28A3 |
| ABCB1 | ATM | CYP2C19 | ERCC1 | MTHFR | SLCO1B3 |
| ABCB4 | ATP7A | CYP2C19 | ERCC2 | NAT | SOD2 |
| ABCC | BRCA1 | CYP2C8 | FANCD2 | NAT2 | STMN1 |
| ABCC1 | CASP3 | CYP2C8*3 | FGD4 | NCF4 | TOP2B |
| ABCC2 | CAT | CYP2C9 | FGFR4 | NOS3 | TP53 |
| ABCC4 | CBR1 | CYP2J2 | FMO3 | NQO1 | TUBB2A |
| ABCG2 | CBR3 | CYP2W1 | GST | NQO2 | TYMP |
| AKR1C3 | CD95 | CYP3A4 | GSTA1 | NT5C2 | TYMS |
| ALDH | CHST3 | CYP3A4*22 | GSTM1 | OMPDC | UGT |
| ALDH1A1 | CRYAB | CYP3A5 | GSTM3 | OPRT | UGT1A1 |
| ALDH3A | CYBA | CYP3A5 | GSTT1 | PXR | UGT1A9 |
| ATG12 | CYP1B1 | CYP3A7 | MAP1LC-38 | RAC2 | UGT2B7 |
| ATG13 | CYP2A13 | DPYD | MBL2 | SLC10A2 | |
| ATG5 | CYP2A6 | EPHA5 | MRP | SLC22A16 | |

### *3.8. Identification of pharmacogenomic variants*

From the 17 TNBC WES annotated file, filtering steps were done separately by matching the annotated file of TNBC patients with PharmGKB gene list and the same with metabolizing gene list. The second filter retained only those variants which are exonic and splice site in function. The next filtering retained only those variants that are predicted to be pathogenic/deleterious by any two of the pathogenicity predicting tools- Sorting Intolerant from Tolerant (SIFT), Polymorphism Phenotyping v2 (PolyPhen v2) and Mutation Taster. The resulting variants were checked using PharmGKB and ClinVar.

### *3.9. Pathway analysis using DAVID*

The resulting variants after the third filtering steps were used for pathway analysis using Database for Annotation, Visualization and Integrated Discovery (DAVID). It offers a wide range of functional annotation tools that helps to reveal the significance of interacting genes. It is based on DAVID Knowledgebase v2023q2, which is dependent on the DAVID Gene concept, drawing various functional annotation sources. The variants were replaced by their respective gene names and the duplicates were removed. A total of 111 genes were selected for pathway analysis. After the gene list was pasted in the gene list qwerty box, "OFFICIALGENESYMBOL" was selected as identifier and "*Homo sapiens*" as species. From the Annotation results further analysis was done using Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway and REACTOME pathway.

### *3.10. Identification of Druggable Genes*

This study also has information regarding to which genes got altered, therefore, identifying the genes that are druggable is of the highest merit. In this study, the TNBC annotated files were converted to .maf files using perl script, this was then used in R-Package as an input file. The Maftools in R-package was used to analyse the MAF file to find the Drug-Interaction Function using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions.

### *3.11. Inter-population differences in the allele frequency of important pharmacogenomic variants*

Variants that passed the filtering steps mentioned above and those present in PharmGKB and ClinVar database were used for this analysis. The variants were identified by their rsID and the allele frequency of the observed variants was retrieved from IndiGenomes database and the same was taken from The 1000 Genomes Project (1000g), The Exome Aggregation Consortium (ExAC), The

Genome Aggregation Database (gnomAD) and Exome Sequencing Project (esp6000). The allele frequency was grouped into five groups as 0.0, 0.25, 0.50, 0.75 and 1.0, after which Fisher Exact test was performed on the allele frequencies and Flourish Studio was used to represent the significant variants graphically.

### *3.12. Establishing common polymorphic variants among TNBC patients*

The WES data of 17 TNBC patients were clubbed together and variants that has allele frequency lower than 0.5 were excluded. From the remaining variants all duplicate variants that were present in different TNBC patients were excluded to remove redundancy of the catalogue variants. ClinVar database was used to check whether these filtered variants have any association with any disease, then, those variants with no disease association were then taken further. Variants that had predicted to be benign in all three pathogenic score such as SIFT, Polyphen2 and Mutation Taster were then filtered out. These variants were then matched with healthy controls and those that were also present in healthy controls were considered to be common polymorphic among mizo population.

### *3.13. Interpretation of population diversity using PCA plot*

Principal component analysis (PCA) (Pearson, 1901) was used to identify the genetic and geographical distances between population by preserving their covariance matrixes. 1000Gphase1.snps vcf and their indexed files were downloaded from the ftp server of GATK Resource Bundle. Linkage disequilibrium pruning of our sample and 1000 genomes was performed using reference genome GRCh38, and the common variants between Mizoram population and 1000 genome data were extracted and merged using bcftools. Eigen vectors were computed using Plink software

## Chapter 4

## Results

### *4.1. Sample details*

About 28 BC patients diagnosed as TNBC and admitted in Mizoram State Cancer Institute from 2018 to 2020 were recruited for the study after obtaining the signed consent form for participating in this study. Their medical records from the time of diagnosis, treatment and ADRs and follow-up questionnaires were collected. Clinical records such as histopathology, tumour grade, Modified Bloom-Richardson (MBR) grade, axillary lymph node involvement, chemotherapeutic treatment administered and information regarding ADR's were also collected using a well-structured questionnaire. To study the genomic alterations and its implications in malignancy and toxicity against the drugs administered, peripheral blood of 5 ml was drawn from each TNBC patients.

Table.3.1. List of Triple Negative Breast Cancer recruited for the study

| SampleID | Sex | Age | DoD |
|---|---|---|---|
| BC6 | F | 91 | 11.06.2018 |
| BC12 | F | 82 | 01.05.2018 |
| BC14 | F | 58 | 01.09.2018 |
| BC26 | F | 41 | 25.10.2018 |
| BC34 | F | 56 | 10.01.2019 |
| BC41 | F | 53 | 15.02.2019 |
| BC44 | F | 34 | 22.02.2019 |
| BC50 | F | 55 | 07.03.2019 |
| BC54 | F | 42 | 00.12.2018 |
| BC56 | F | 70 | 20.04.2019 |
| BC62 | F | 54 | 23.03.2019 |
| BC67 | F | 55 | 08.06.2019 |
| BC71 | F | 50 | 10.06.2019 |
| BC75 | F | 53 | 04.07.2019 |

| | | | |
|---|---|---|---|
| BC77 | F | 28 | 16.07.2019 |
| BC86 | F | 48 | 07.08.2019 |
| BC89 | F | 63 | 31.07.2019 |
| BC91 | F | 40 | 09.09.2019 |
| BC102 | F | 51 | 14.10.2019 |
| BC112 | F | 66 | 19.10.2019 |
| BC113 | F | 55 | 06.11.2019 |
| BC146 | F | 57 | 00.06.2019 |
| BC147 | F | 36 | 17.06.2020 |
| BC148 | F | 47 | 16.06.2020 |
| BC150 | F | 50 | 01.06.2020 |
| BC156 | F | 49 | 18.07.2020 |
| BC157 | F | 44 | 01.07.2020 |
| BC163 | F | 32 | 29.07.2020 |

*DoD: Date of Diagnosis*

### *4.2. Representation of DNA Agarose Gel Electrophoresis picture*

To study the genomic alterations and its implications in malignancy and toxicity against the drugs administered to the patients, peripheral blood of 5 ml was drawn from each TNBC patients. DNA was isolated using QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN) as per the manufacturer's protocol with little modifications and the eluted DNA was visualised using UV Gel documentation system (Bio-RAD) and representation of genomic DNA isolates are given in figure1

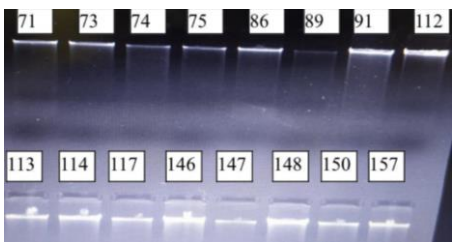



Figure. 4.2.: Triple Negative Breast Cancer Genomic DNA gel picture of representative samples

### *4.3. List of ADRs observed in TNBC patients*

For ADR information the National Comprehensive Cancer Network (NCCN) guidelines was followed and list of ADR observed is listed in Table .4.3

Table 4.3.: List of ADRs observed in TNBC patients

| List of ADRs | No of Patients |
|---|---|
| Alopecia | 25 |
| Anorexia | 9 |
| Anxiety | 2 |
| Arthritis | 1 |
| Auditory | 1 |
| Blurred vision | 6 |
| Cardiac Arrhythmia | 3 |
| Change in appetite | 15 |
| Constipation | 13 |
| Darkening of nails | 10 |
| Dehydration | 2 |
| Diabetes | 6 |
| Diarrhoea | 5 |
| Drowsiness | 18 |
| Dry skin | 1 |
| Fatigue | 8 |
| Gastritis | 4 |
| Heartburn | 5 |
| Itching | 3 |
| Leukopenia | 9 |
| Mood alteration | 4 |
| Mucositis-Stomach | 1 |
| Nausea | 21 |
| Nephritis | 3 |
| Oral cavity | 1 |
| Oral ulceration | 1 |
| Pneumonitis | 2 |
| Swelling | 1 |
| Urinary tract infection | 1 |
| Vomiting | 7 |

### *4.4. ADRs observed in TNBC patients with their severity*

From the TNBC patients, the questionnaire was structured to identify the severity of the ADR observed into mild, moderate and severe and are listed in Table 3

Table 4.4.: ADRs observed in TNBC patients with their severity

| Sl. No | ADRs | Severity | | | No of Patients |
|---|---|---|---|---|---|
| | | Mild (%) | Moderate (%) | Severe (%) | |
| 1 | Alopecia | 0 (0) | 1 (4) | 24 (96) | 25 (89.28) |
| 2 | Anorexia | 1 (11.11) | 8 (88.88) | 0 (0) | 9 (32.14) |
| 3 | Anxiety | 2 (100) | 0 (0) | 0 (0) | 2 (7.14) |
| 4 | Arthritis | 0 (0) | 0 (0) | 1 (100) | 1 (3.57) |
| 5 | Auditory | 1 (100) | 0 (0) | 0 (0) | 1 (3.57) |
| 6 | Blurred vision | 6 (100) | 0 (0) | 0 (0) | 6 (21.42) |
| 7 | Cardiac Arrhythmia | 3 (75) | 0 (0) | 1 (25) | 4 (14.28) |
| 8 | Change in appetite | 5 (29.41) | 11 (64.70) | 1 (5.88) | 17 (60.71) |
| 9 | Constipation | 2 (15.38) | 6 (46.15) | 5 (38.46) | 13 (46.42) |
| 10 | Darkening of nails | 3 (30) | 7 (70) | 0 (0) | 10 (35.71) |
| 11 | Dehydration | 0 (0) | 2 (100) | 0 (0) | 2 (7.14) |
| 12 | Diabetes | 0 (0) | 4 (66.66) | 2 (33.33) | 6 (21.42) |
| 13 | Diarrhoea | 2 (40) | 2 (40) | 1 (20) | 5 (17.85) |

| | | | | | |
|---|---|---|---|---|---|
| 14 | Drowsiness | 3 (14.28) | 13 (61.904) | 5 (23.80) | 21 (75) |
| 15 | Dry skin | 0 (0) | 1 (100) | 0 (0) | 1 (3.57) |
| 16 | Fatigue | 1 (12.5) | 6 (75) | 1 (12.5) | 8 (28.57) |
| 17 | Gastritis | 1 (25) | 3 (75) | 0 (0) | 4 (14.28) |
| 18 | Heartburn | 3 (60) | 2 (40) | 0 (0) | 5 (17.85) |
| 19 | Itching | 0 (0) | 2 (66.66) | 1 (33.33) | 3 (10.71) |
| 20 | Leukopenia | 2 (22.22) | 6 (66.66) | 1 (11.11) | 9 (32.14) |
| 21 | Mood alteration | 4 (100) | 0 (0) | 0 (0) | 4 (14.28) |
| 22 | Mucositis-Stomach | 0 (0) | 0 (0) | 1 (100) | 1 (3.57) |
| 23 | Nausea | 6 (28.57) | 12 (57.142) | 3 (14.28) | 21 (75) |
| 24 | Nephritis | 2 (66.66) | 1 (33.33) | 0 (0) | 3 (10.71) |
| 25 | Oral cavity | 1 (100) | 0 (0) | 0 (0) | 1 (3.57) |
| 26 | Oral ulceration | 0 (0) | 1 (100) | 0 (0) | 1 (3.57) |
| 27 | Pneumonitis | 1 (50) | 1 (50) | 0 (0) | 2 (7.14) |
| 28 | Swelling | 0 (0) | 1 (100) | 0 (0) | 1 (3.57) |
| 29 | Urinary tract infection | 0 (0) | 1 (100) | 0 (0) | 1 (3.57) |
| 30 | Vomiting | 2 (28.57) | 5 (71.42) | 0 (0) | 7 (25) |

### *4.5. Total exonic variants observed after filtering*

WES was performed to understand the genomic background of TNBC patients that might play a role in clinical settings by using PharmGKB database, DrugBank and through literature search.

Table. 4.7.: List of total exonic variants observed after filtering (n = 125)

| **Chr** | **Start** | **End** | **Ref** | **Alt** | **Gene** | **ExonicFunct.** | **rsID** |
|---|---|---|---|---|---|---|---|
| chr1 | 156844494 | 156844494 | G | T | INSRR | Non-synonymous | rs374549328 |
| chr1 | 158620352 | 158620352 | G | A | SPTA1 | | rs372075053 |
| chr11 | 103155401 | 103155401 | G | A | DYNC2H1 | | . |
| chr16 | 2279097 | 2279097 | C | T | ABCA3 | | rs201955122 |
| chr16 | 72023522 | 72023522 | C | T | DHODH | | rs61733129 |
| chr3 | 113379370 | 113379370 | G | A | CFAP44 | | rs58988763 |
| chr3 | 194361024 | 194361024 | A | T | LRRC15 | | rs149554510 |
| chr5 | 180603278 | 180603278 | C | T | FLT4 | | rs201796032 |
| chr6 | 160134079 | 160134079 | C | T | SLC22A1 | stopgain | . |
| chr6 | 160136611 | 160136611 | C | T | SLC22A1 | Non-synonymous | rs2282143 |
| chr1 | 67182913 | 67182913 | G | A | IL23R | | rs76418789 |
| chr10 | 52771482 | 52771482 | G | A | MBL2 | | rs5030737 |
| chr12 | 122940277 | 122940277 | G | C | ABCB9 | | rs116062676 |
| chr17 | 65541552 | 65541552 | C | T | AXIN2 | | . |
| chr19 | 35848142 | 35848142 | C | T | NPHS1 | | rs28939695 |
| chr3 | 7578746 | 7578746 | C | T | GRM7 | | rs141874818 |
| chr3 | 97610853 | 97610853 | G | A | EPHA6 | | rs776617495 |
| chr5 | 150117817 | 150117817 | T | G | PDGFRB | | rs879255377 |
| chr6 | 121223292 | 121223292 | G | A | TBC1D32 | stopgain | . |
| chr7 | 140020029 | 140020029 | C | T | TBXAS1 | Non-synonymous | rs13306050 |
| chr1 | 33533855 | 33533855 | G | A | CSMD2 | | rs148658404 |
| chr10 | 60173110 | 60173110 | G | A | ANK3 | | rs190581397 |
| chr2 | 205723790 | 205723790 | C | A | NRP2 | | rs532796522 |
| chr2 | 233810833 | 233810833 | C | T | MROH2A | stopgain | rs28900672 |
| chr20 | 59022915 | 59022915 | A | C | TUBB1 | | rs463312 |
| chr6 | 41061590 | 41061590 | G | A | APOBEC2 | | rs142866037 |
| chr7 | 116699722 | 116699722 | C | T | MET | | rs367722737 |
| chr7 | 150856754 | 150856754 | G | A | AOC1 | | rs572549369 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr1 | 26338416 | 26338416 | A | C | CRYBG2 | Non-synonymous | rs151324745 |
| chr15 | 92946064 | 92946064 | A | C | CHD2 | | rs531251618 |
| chr17 | 68907860 | 68907860 | A | C | ABCA8 | | rs16973424 |
| chr6 | 24848144 | 24848144 | C | T | RIPOR2 | | rs35331811 |
| chr6 | 43770891 | 43770891 | C | T | VEGFA | | . |
| chr8 | 132141191 | 132141191 | T | C | KCNQ3 | | rs118192252 |
| chr1 | 156243142 | 156243142 | C | T | BGLAP | | rs200610507 |
| chr16 | 56902407 | 56902407 | C | T | SLC12A3 | | rs12708965 |
| chr17 | 43063937 | 43063937 | A | G | BRCA1 | | rs80356993 |
| chr17 | 50667565 | 50667565 | G | T | ABCC3 | | . |
| chr2 | 21029662 | 21029662 | G | A | APOB | | rs13306194 |
| chr3 | 151387919 | 151387919 | G | A | MED12L | | rs2276761 |
| chr6 | 52237046 | 52237046 | T | C | IL17F | | rs2397084 |
| chr6 | 133483033 | 133483033 | G | A | EYA4 | | rs143936434 |
| chr6 | 137874854 | 137874854 | A | G | TNFAIP3 | | rs146534657 |
| chr6 | 138092103 | 138092103 | G | A | PERP | | rs75183345 |
| chr13 | 51949772 | 51949772 | G | A | ATP7B | | rs121907993 |
| chr16 | 28593476 | 28593476 | C | A | SULT1A2 | | . |
| chr2 | 21012072 | 21012072 | C | T | APOB | | rs746414462 |
| chr22 | 24321443 | 24321443 | C | T | SPECC1L | | rs148203655 |
| chr4 | 186082866 | 186082866 | C | T | TLR3 | stopgain | rs780744847 |
| chr1 | 33625215 | 33625215 | G | A | CSMD2 | Non-synonymous | . |
| chr1 | 201066950 | 201066950 | C | T | CACNA1S | | rs545411173 |
| chr12 | 21300535 | 21300535 | C | T | SLCO1A2 | | rs754967522 |
| chr15 | 81299420 | 81299420 | G | T | IL16 | | . |
| chr2 | 159453669 | 159453669 | C | T | BAZ2B | | rs766155088 |
| chr2 | 167246619 | 167246619 | G | A | XIRP2 | | rs181539061 |
| chr3 | 57446584 | 57446584 | C | T | DNAH12 | | rs73076451 |
| chr4 | 102612562 | 102612562 | C | T | NFKB1 | | rs747643116 |
| chr10 | 94687805 | 94687805 | T | A | CYP2C18 | stopgain | rs41291550 |
| chr10 | 94780653 | 94780653 | G | A | CYP2C19 | | rs4986893 |
| chr10 | 97057271 | 97057271 | C | T | SLIT1 | Non-synonymous | . |
| chr12 | 71577986 | 71577986 | C | G | LGR5 | | rs138215571 |
| chr12 | 102877479 | 102877479 | G | C | PAH | | . |
| chr18 | 6824836 | 6824836 | G | A | ARHGAP28 | | rs188975691 |
| chr5 | 96800974 | 96800974 | G | C | ERAP1 | | rs142482678 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr7 | 151009457 | 151009457 | C | T | NOS3 | | rs368180942 |
| chr1 | 46408574 | 46408574 | C | T | FAAH | | rs77101686 |
| chr1 | 113726481 | 113726481 | T | C | PHTF1 | | . |
| chr12 | 69697672 | 69697672 | A | G | BEST3 | | rs1025016 |
| chr15 | 51222502 | 51222502 | G | A | CYP19A1 | | rs201842322 |
| chr2 | 233032793 | 233032793 | G | A | NEU2 | | rs2233385 |
| chr4 | 39447442 | 39447442 | T | C | KLB | | rs17618262 |
| chr8 | 104451318 | 104451318 | C | T | DPYS | stopgain | . |
| chr1 | 97305364 | 97305364 | C | T | DPYD | | rs1801160 |
| chr11 | 75005458 | 75005458 | C | T | NEU3 | | rs199990681 |
| chr16 | 1200281 | 1200281 | C | T | CACNA1H | | rs201911360 |
| chr17 | 45112932 | 45112932 | C | T | PLCD3 | | . |
| chr3 | 133757854 | 133757854 | A | G | TF | | rs41295774 |
| chr6 | 18130762 | 18130762 | C | T | TPMT | | rs56161402 |
| chr19 | 44908822 | 44908822 | C | T | APOE | | rs7412 |
| chr4 | 65602256 | 65602256 | A | G | EPHA5 | | . |
| chr5 | 635393 | 635393 | C | T | CEP72 | | rs869955 |
| chr13 | 35670933 | 35670933 | T | C | NBEA | | . |
| chr19 | 45778500 | 45778500 | C | T | DMPK | Non-synonymous | rs145330026 |
| chr3 | 57446218 | 57446218 | T | C | DNAH12 | | rs62622492 |
| chr5 | 90685792 | 90685792 | C | G | ADGRV1 | | . |
| chr9 | 74785884 | 74785884 | A | G | TRPM6 | | . |
| chr9 | 122381448 | 122381448 | G | A | PTGS1 | | rs201523045 |
| chr21 | 40080175 | 40080175 | A | G | DSCAM | | rs770311896 |
| chr4 | 6301752 | 6301752 | C | T | WFS1 | | rs201064551 |
| chr5 | 178986522 | 178986522 | G | A | GRM6 | | rs62638210 |
| chr8 | 37965844 | 37965844 | G | C | ADRB3 | | . |
| chr1 | 169514323 | 169514323 | T | C | F5 | | rs6027 |
| chr11 | 113207837 | 113207837 | G | A | NCAM1 | | . |
| chr12 | 71577978 | 71577978 | T | C | LGR5 | | rs753804853 |
| chr19 | 45779470 | 45779470 | A | G | DMPK | | rs562015011 |
| chr2 | 127292756 | 127292756 | G | A | ERCC3 | stopgain | rs34295337 |
| chr2 | 214989581 | 214989581 | T | C | ABCA12 | | . |
| chr3 | 121922803 | 121922803 | G | T | SLC15A2 | Non-synonymous | rs376349271 |
| chr4 | 2992236 | 2992236 | G | C | GRK4 | | rs13305979 |
| chr7 | 130275233 | 130275233 | T | G | CPA2 | | rs145868138 |

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| chr16 | 8764825 | 8764825 | T | A | ABAT | | . |
| chr3 | 57542834 | 57542834 | T | C | DNAH12 | | . |
| chr4 | 186620329 | 186620329 | T | C | FAT1 | | . |
| chr10 | 95043067 | 95043067 | C | G | CYP2C8 | | rs529746836 |
| chr11 | 113412762 | 113412762 | G | C | DRD2 | | rs1801028 |
| chr1 | 11796321 | 11796321 | G | A | MTHFR | | rs1801133 |
| chr1 | 169560616 | 169560616 | T | C | F5 | | rs201510575 |
| chr1 | 201040054 | 201040054 | A | G | CACNA1S | | rs12139527 |
| chr12 | 21178615 | 21178615 | T | C | SLCO1B1 | | rs4149056 |
| chr12 | 47879112 | 47879112 | A | G | VDR | startloss | rs2228570 |
| chr13 | 48037798 | 48037798 | G | A | NUDT15 | Non- | rs186364861 |
| chr16 | 28620062 | 28620062 | C | G | SULT1A1 | synonymous. | rs79527462 |
| chr17 | 43104187 | 43104187 | G | A | BRCA1 | stopgain | rs1085307902 |
| chr17 | 63479068 | 63479068 | A | G | ACE | | rs117134739 |
| chr17 | 63483970 | 63483970 | C | T | ACE | | rs567828872 |
| chr19 | 15879621 | 15879621 | C | T | CYP4F2 | | rs2108622 |
| chr19 | 38469316 | 38469316 | G | A | RYR1 | | rs749283427 |
| chr19 | 38477794 | 38477794 | G | A | RYR1 | | rs753735721 |
| chr19 | 38485972 | 38485972 | C | T | RYR1 | Non- | rs192863857 |
| chr19 | 38494546 | 38494546 | G | A | RYR1 | synonymous | rs554066182 |
| chr19 | 40991369 | 40991369 | C | T | CYP2B6 | | rs8192709 |
| chr3 | 38575385 | 38575385 | C | T | SCN5A | | rs41261344 |
| chr4 | 88131113 | 88131113 | C | T | ABCG2 | | rs528655917 |
| chr5 | 148827037 | 148827037 | A | G | ADRB2 | | rs201257377 |
| chrX | 154534495 | 154534495 | C | T | G6PD | | rs137852314 |

### *4.6. Distribution of variants in PharmGKB database, ClinVar and dbSNP*

The observed 125 variants were then plotted in VENN Diagram to understand which variants which are overlapping in PharmGKB database, ClinVar (significant in ClinVar and Describe for other disease), dbSNP and not reported in dbSNP (Figure. 4.8.).

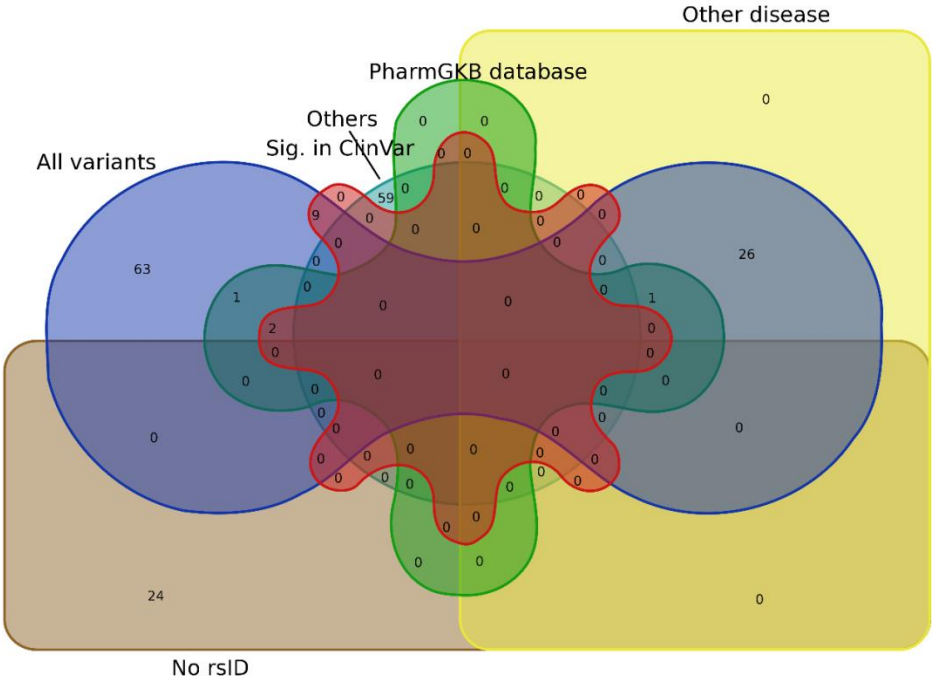



Figure. 4.8.: Distribution of variants observed after filtering

### *4.9. Variants that are present in PharmGKB Database*

After undergoing hard filtering steps, we observed four variants that were showing relations to toxicity when treated with chemotherapeutic drugs in MSCI, Zemabawk with their level of evidence (Table 4.9.)

Table. 4.9.: Variants that are present in PharmGKB Database with their effects

| | | | | |
|---|---|---|---|---|
| **Chr** | chr1 | chr1 | chr12 | chr19 |
| **Start** | 97305364 | 11796321 | 21178615 | 40991369 |
| **End** | 97305364 | 11796321 | 21178615 | 40991369 |
| **Ref** | C | G | T | C |
| **Alt** | T | A | C | T |
| **Gene** | *DPYD* | *MTHFR* | *SLCO1B1* | *CYP2B6* |
| **PharmGKB ID** | 1.45E+09 | 9.81E+08 | 1.45E+09 | 6.55E+08 |
| **Level** | 1A | 4 | 3 | 3 |
| **Drugs** | fluorouracil | capecitabine; fluorouracil | cyclophosphamide; docetaxel; doxorubicin; epirubicin; fluorouracil | cyclophosphamide |
| **Effects** | Toxicity | Toxicity | Toxicity | Toxicity |
| **Phenotype** | Neoplasms | Neoplasms | Breast Neoplasms | Cystitis; Transplantation |
| **AF in 1000g** | 0.04393 | 0.245407 | 0.08766 | 0.047524 |

### *4.10. Variants present in ClinVar as pathogenic*

Those variants which were not present in PharmGKB datase were then overlapped with variants present in ClinVar that showed Pathogenicity in their clincal significance. We observed 11 varaints and is depicted in Figure. 4.10. with number of patients who harbour the variants.

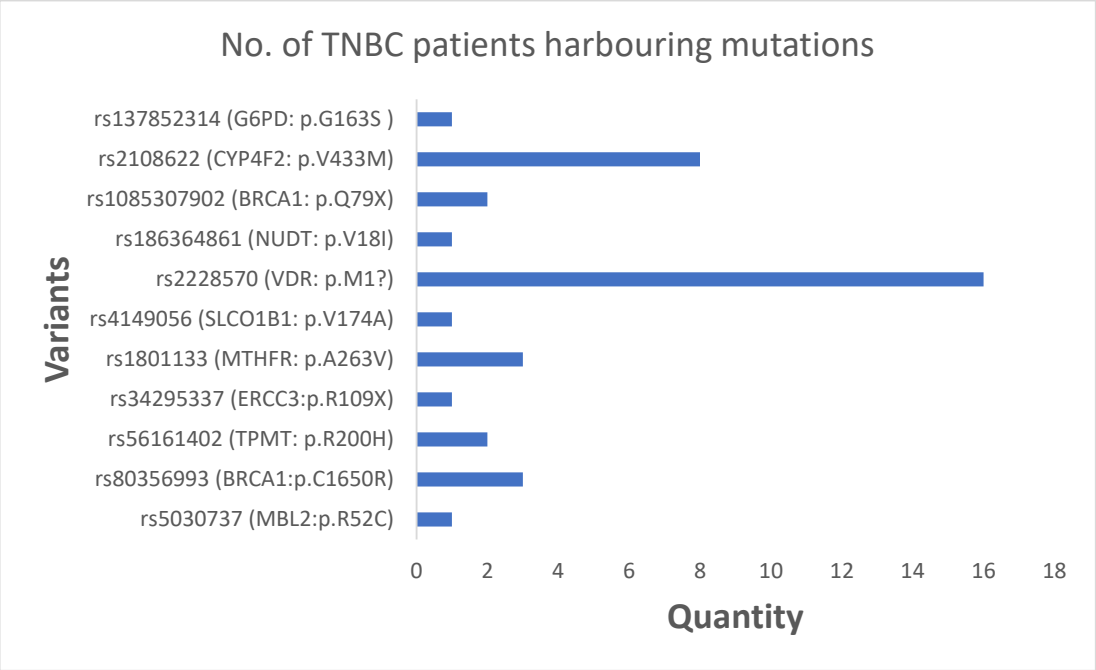
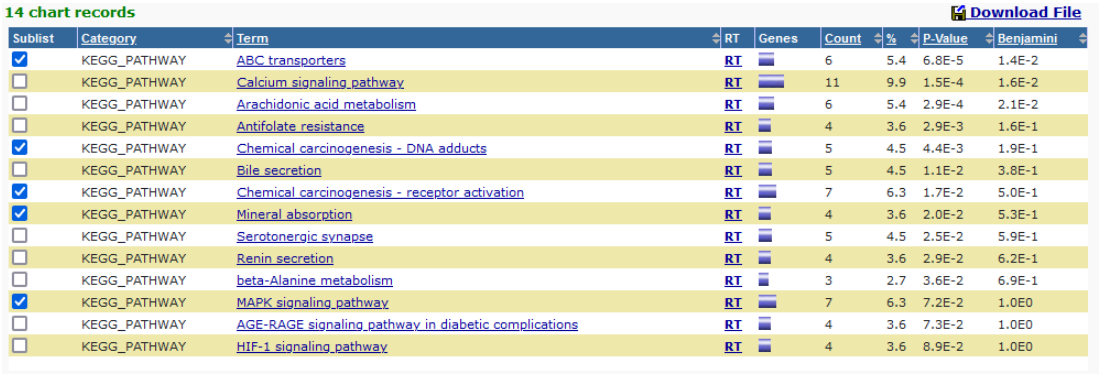



Figure. 4.10.: Variants thar are present in ClinVar with Clinical Significance as Pathogenic and Drug Response

### *4.11. Pathway analysis using DAVID (Kegg pathway)*

From the WES data (annotated file), all the genes were clubbed from each of the seventeen TNBC patients and the duplicated genes were removed. This list of genes was used for pathway analysis using DAVID with the latest DAVID Knowledgebase v2023q2. Figure.4.11. shows the list of 14 pathways obtained using KEGG database.



14 chart records — Download File

| Sublist | Category | Term | RT | Genes | Count | % | P-Value | Benjamini |
|---|---|---|---|---|---|---|---|---|
| ☑ | KEGG_PATHWAY | ABC transporters | RT | | 6 | 5.4 | 6.8E-5 | 1.4E-2 |
| ☐ | KEGG_PATHWAY | Calcium signaling pathway | RT | | 11 | 9.9 | 1.5E-4 | 1.6E-2 |
| ☐ | KEGG_PATHWAY | Arachidonic acid metabolism | RT | | 6 | 5.4 | 2.9E-4 | 2.1E-2 |
| ☐ | KEGG_PATHWAY | Antifolate resistance | RT | | 4 | 3.6 | 2.9E-3 | 1.6E-1 |
| ☑ | KEGG_PATHWAY | Chemical carcinogenesis - DNA adducts | RT | | 5 | 4.5 | 4.4E-3 | 1.9E-1 |
| ☐ | KEGG_PATHWAY | Bile secretion | RT | | 5 | 4.5 | 1.1E-2 | 3.8E-1 |
| ☑ | KEGG_PATHWAY | Chemical carcinogenesis - receptor activation | RT | | 7 | 6.3 | 1.7E-2 | 5.0E-1 |
| ☑ | KEGG_PATHWAY | Mineral absorption | RT | | 4 | 3.6 | 2.0E-2 | 5.3E-1 |
| ☐ | KEGG_PATHWAY | Serotonergic synapse | RT | | 5 | 4.5 | 2.5E-2 | 5.9E-1 |
| ☐ | KEGG_PATHWAY | Renin secretion | RT | | 4 | 3.6 | 2.9E-2 | 6.2E-1 |
| ☐ | KEGG_PATHWAY | beta-Alanine metabolism | RT | | 3 | 2.7 | 3.6E-2 | 6.9E-1 |
| ☑ | KEGG_PATHWAY | MAPK signaling pathway | RT | | 7 | 6.3 | 7.2E-2 | 1.0E0 |
| ☐ | KEGG_PATHWAY | AGE-RAGE signaling pathway in diabetic complications | RT | | 4 | 3.6 | 7.3E-2 | 1.0E0 |
| ☐ | KEGG_PATHWAY | HIF-1 signaling pathway | RT | | 4 | 3.6 | 8.9E-2 | 1.0E0 |

Figure.4.11.: Pathway analysis using DAVID (KEGG pathway)

### *4.12. Pathway analysis using DAVID (Reactome)*

Similarly, we observed 45 pathways using Reactome database.

45 chart records — Download File

| Sublist | Category | Term | RT | Genes | Count | % | P-Value | Benjamini |
|---|---|---|---|---|---|---|---|---|
| ☑ | REACTOME_PATHWAY | Drug ADME | RT | | 10 | 9.0 | 3.9E-7 | 2.3E-4 |
| ☐ | REACTOME_PATHWAY | Biological oxidations | RT | | 12 | 10.8 | 2.9E-6 | 6.4E-4 |
| ☐ | REACTOME_PATHWAY | Phase I - Functionalization of compounds | RT | | 9 | 8.1 | 3.3E-6 | 6.4E-4 |
| ☐ | REACTOME_PATHWAY | Cytochrome P450 - arranged by substrate type | RT | | 7 | 6.3 | 2.0E-5 | 3.0E-3 |
| ☑ | REACTOME_PATHWAY | Transport of small molecules | RT | | 18 | 16.2 | 1.3E-4 | 1.4E-2 |
| ☐ | REACTOME_PATHWAY | Arachidonic acid metabolism | RT | | 6 | 5.4 | 1.4E-4 | 1.4E-2 |
| ☐ | REACTOME_PATHWAY | Nervous system development | RT | | 15 | 13.5 | 3.9E-4 | 3.2E-2 |
| ☐ | REACTOME_PATHWAY | L1CAM interactions | RT | | 7 | 6.3 | 5.4E-4 | 3.9E-2 |
| ☐ | REACTOME_PATHWAY | Ciprofloxacin ADME | RT | | 3 | 2.7 | 7.2E-4 | 4.6E-2 |
| ☐ | REACTOME_PATHWAY | Axon guidance | RT | | 14 | 12.6 | 8.4E-4 | 4.9E-2 |
| ☑ | REACTOME_PATHWAY | Xenobiotics | RT | | 4 | 3.6 | 1.2E-3 | 6.5E-2 |
| ☐ | REACTOME_PATHWAY | Interaction between L1 and Ankyrins | RT | | 4 | 3.6 | 2.3E-3 | 1.1E-1 |
| ☑ | REACTOME_PATHWAY | Metabolism | RT | | 31 | 27.9 | 2.5E-3 | 1.1E-1 |
| ☐ | REACTOME_PATHWAY | Synthesis of (16-20)-hydroxyeicosatetraenoic acids (HETE) | RT | | 3 | 2.7 | 2.5E-3 | 1.1E-1 |
| ☑ | REACTOME_PATHWAY | CYP2E1 reactions | RT | | 3 | 2.7 | 3.8E-3 | 1.5E-1 |
| ☐ | REACTOME_PATHWAY | Recycling of bile acids and salts | RT | | 3 | 2.7 | 1.0E-2 | 3.7E-1 |
| ☑ | REACTOME_PATHWAY | ABC-family proteins mediated transport | RT | | 5 | 4.5 | 1.2E-2 | 4.0E-1 |
| ☐ | REACTOME_PATHWAY | Post-translational protein phosphorylation | RT | | 5 | 4.5 | 1.4E-2 | 4.5E-1 |
| ☐ | REACTOME_PATHWAY | NCAM signaling for neurite out-growth | RT | | 4 | 3.6 | 1.7E-2 | 5.2E-1 |
| ☐ | REACTOME_PATHWAY | Fatty acid metabolism | RT | | 6 | 5.4 | 1.8E-2 | 5.3E-1 |
| ☐ | REACTOME_PATHWAY | Regulation of Insulin-like Growth Factor (IGF) transport and uptake by Insulin-like Growth Factor Binding Proteins (IGFBPs) | RT | | 5 | 4.5 | 2.2E-2 | 6.2E-1 |
| ☐ | REACTOME_PATHWAY | Metabolism of lipids | RT | | 13 | 11.7 | 2.5E-2 | 6.5E-1 |
| ☐ | REACTOME_PATHWAY | Paracetamol ADME | RT | | 3 | 2.7 | 2.5E-2 | 6.5E-1 |
| ☑ | REACTOME_PATHWAY | Metabolic disorders of biological oxidation enzymes | RT | | 3 | 2.7 | 3.4E-2 | 8.3E-1 |
| ☐ | REACTOME_PATHWAY | Nucleotide catabolism | RT | | 3 | 2.7 | 3.8E-2 | 8.7E-1 |
| ☐ | REACTOME_PATHWAY | Developmental Biology | RT | | 17 | 15.3 | 3.9E-2 | 8.7E-1 |
| ☐ | REACTOME_PATHWAY | Abacavir transmembrane transport | RT | | 2 | 1.8 | 4.2E-2 | 8.7E-1 |
| ☐ | REACTOME_PATHWAY | Chylomicron clearance | RT | | 2 | 1.8 | 4.2E-2 | 8.7E-1 |
| ☐ | REACTOME_PATHWAY | Metabolism of steroids | RT | | 5 | 4.5 | 4.3E-2 | 8.7E-1 |
| ☐ | REACTOME_PATHWAY | Amine ligand-binding receptors | RT | | 3 | 2.7 | 5.0E-2 | 9.5E-1 |
| ☐ | REACTOME_PATHWAY | NCAM1 interactions | RT | | 3 | 2.7 | 5.0E-2 | 9.5E-1 |
| ☐ | REACTOME_PATHWAY | Metabolism of nucleotides | RT | | 4 | 3.6 | 5.3E-2 | 9.7E-1 |
| ☐ | REACTOME_PATHWAY | Aspirin ADME | RT | | 3 | 2.7 | 5.5E-2 | 9.7E-1 |
| ☐ | REACTOME_PATHWAY | Bile acid and bile salt metabolism | RT | | 3 | 2.7 | 5.7E-2 | 9.8E-1 |
| ☐ | REACTOME_PATHWAY | Signaling by VEGF | RT | | 4 | 3.6 | 6.6E-2 | 1.0E0 |
| ☐ | REACTOME_PATHWAY | Synthesis of epoxy (EET) and dihydroxyeicosatrienoic acids (DHET) | RT | | 2 | 1.8 | 6.7E-2 | 1.0E0 |

Figure. 4.12.: Pathway analysis using DAVID (Reactome)

### *4.13. BRCA1: p.C1650R*

Out of the 11 variants observed from ClinVar indicating Pathogenicity, *BRCA1* variant (p.C1650R) is highlighted as a representation in Figure. 4.13.

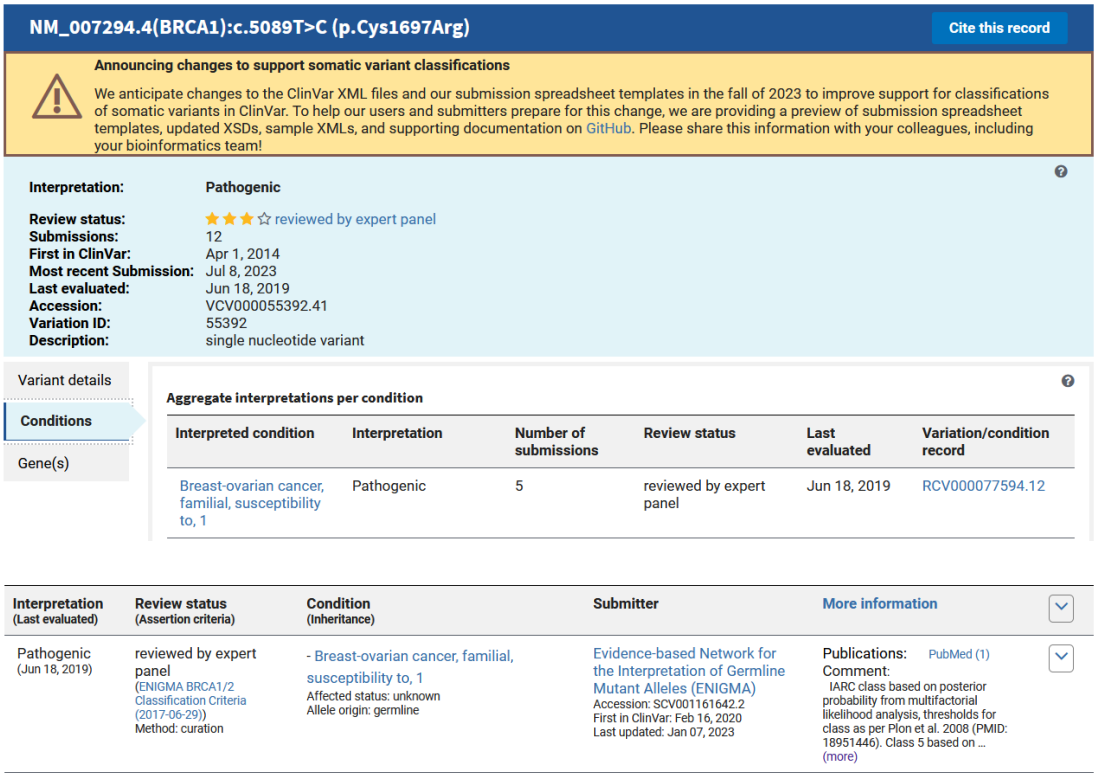



Figure. 4.13.: Representation of one variant (*BRCA1*: p.C1650R) reported to be pathogenic in ClinVar

### *4.14. DPYD (rs1801160), MTHFR (rs1801133), SLCO1B1 (rs4149056) and CYP2B6 (rs8192709)*

We identify 4 variants that were having toxic effects with chemotherapeutic drugs and were present in PharmGKB database. Dihydropyrimidine dehydrogenase (*DPYD*) (p.V732I,) showed toxic effect to patients that were treated with fluorouracil with level of evidence score as 1A. Methylenetetrahydrofolate reductase (*MTHFR*) (p.A222V, rs1801133) showed toxic effect to patients that were treated with fluorouracil and capecitabine with level of evidence score as 4. Solute Carrier organic anion transporter family member 1B1 (*SLCO1B1*) (p.V174A, rs4149056) showed toxic effect to patients that were treated with cyclophosphamide; docetaxel; doxorubicin; epirubicin; fluorouracil with level of evidence score as 3. Cytochrome P450 Family 2 Subfamily B Member 6 (*CYP2B6*) (p.R22C, rs8192709) showed toxic effect to patients that were treated with cyclophosphamide with level of evidence score as 3.

### *4.15. Variants present in ClinVar but has no to little effect*

We observed 27 list of variants that were found in ClinVar but not showed any clinical significance.

Table. 4.15.: Representation of variants that were present in ClinVar but has no to little effect in disease conditions.

| **Gene** | **CLNDN** | **CLNSIG** |
|---|---|---|
| *SPTA1* | Hereditary pyropoikilocytosis | Uncertain significance |
| *ABCA3* | Surfactant metabolism dysfunction | Uncertain significance |
| *DHODH* | Miller syndrome | Benign |
| *CHD2* | Epileptic encephalopathy | Likely benign |
| *KCNQ3* | Benign familial neonatal seizures | Conflicting interpretations of pathogenicity |
| *SLC12A3* | Familial hypokalemia-hypomagnesemia | Benign |
| *APOB* | Warfarin response | Conflicting interpretations of pathogenicity |
| *IL17F* | Candidiasis | Benign |
| *EYA4* | Deafness | Conflicting interpretations of pathogenicity |
| *TNFAIP3* | Autoinflammatory syndrome | Benign |
| *ATP7B* | Wilsondisease | Conflicting interpretations of pathogenicity |
| *APOB* | Familial hypercholesterolemia | Conflicting interpretations of pathogenicity |
| *CACNA1S* | Malignant hyperthermia | Benign/Likely benign |

| | | |
|---|---|---|
| *CYP19A1* | Aromatasedeficiency | Benign |
| *DPYD* | Dihydropyrimidine dehydrogenase deficiency | Benign/Likely benign |
| *TF* | Atransferrinemia | Benign |
| *WFS1* | Type 2 diabetes mellitus | Uncertain significance |
| *F5* | Thrombophilia due to factor V Leiden | Benign/Likely benign |
| *DRD2* | Dystonia | Benign |
| *F5* | Budd-Chiari syndrome | Conflicting interpretations of pathogenicity |
| *CACNA1S* | Hypokalemic periodic paralysis 1 | Benign |
| *ACE* | Renal dysplasia | Benign |
| *RYR1* | RYR1-Related Disorders | Uncertain significance |
| *RYR1* | RYR1-Related Disorders | Uncertain significance |
| *RYR1* | Central core myopathy | Conflicting interpretations of pathogenicity |
| *RYR1* | RYR1-Related Disorders | Uncertain significance |
| *SCN5A* | Arrhythmia | Benign/Likely benign |

### *4.16. Druggable categories and Drug Target*

WES annotated data from all the patients were clubbed, after removing the duplicates, using Maftools in R-package, Drug-Interaction Function was identified using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions. A total of 11 druggable categories and 11 druggable genes were identified.

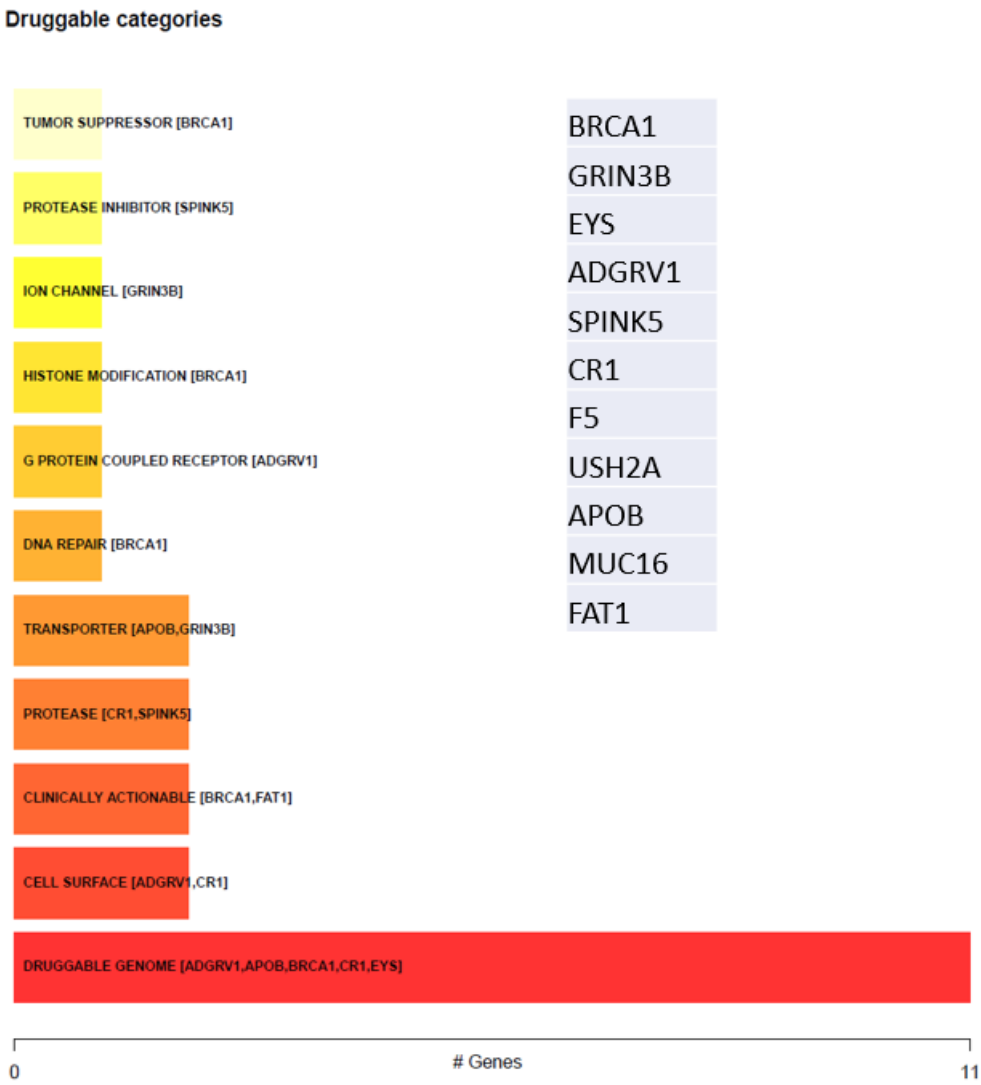



Figure. 4.16.: List of Druggable Categories and Druggable target obtained by Drug Interaction function using Drug Gene Interaction database.

### *4.17. Druggable targets and its Drugs*

As a representation of the identified Druggable targets and its Drugs are listed in

Table. 4.16.: Drug–gene interactions observed from Drug Gene Interaction database

| **Gene** | **Interactiontypes** | **Drugname** | **Drugclaimname** |
|---|---|---|---|
| *BRCA1* | | Doxorubicin | Doxorubicin |
| *BRCA1* | | Veliparib | Veliparib |
| *BRCA1* | | Cyclophosphamide | Cyclophosphamide |
| *BRCA1* | | Olaparib | Olaparib |
| *BRCA1* | | Carboplatin | Carboplatin |
| *BRCA1* | | Tamoxifen | Tamoxifen |
| *BRCA1* | | Paclitaxel | Paclitaxel |
| *MUC16* | | Docetaxel | Docetaxel |
| *F5* | | Tamoxifen | Tamoxifen |

### *4.18. Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1*

Targeted panel sequencing of *BRCA1*, *BRCA2*, *CHEK2* and *PALB1* reveals 39 variants with their frequency of prevalence among TNBC patients (counts).

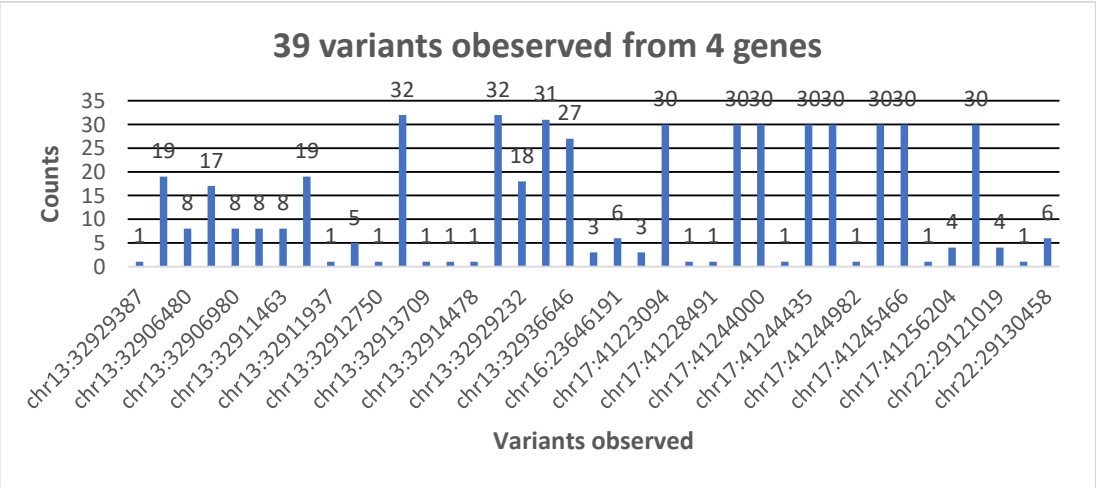



Figure. 4.18a.: Targeted Panel Gene Sequencing of *BRCA1*, *BRCA2*, *CHEK2* and *PALB2* identifies 39 variants.

The thirty-nine variants were classified according to the location they reside within the genes

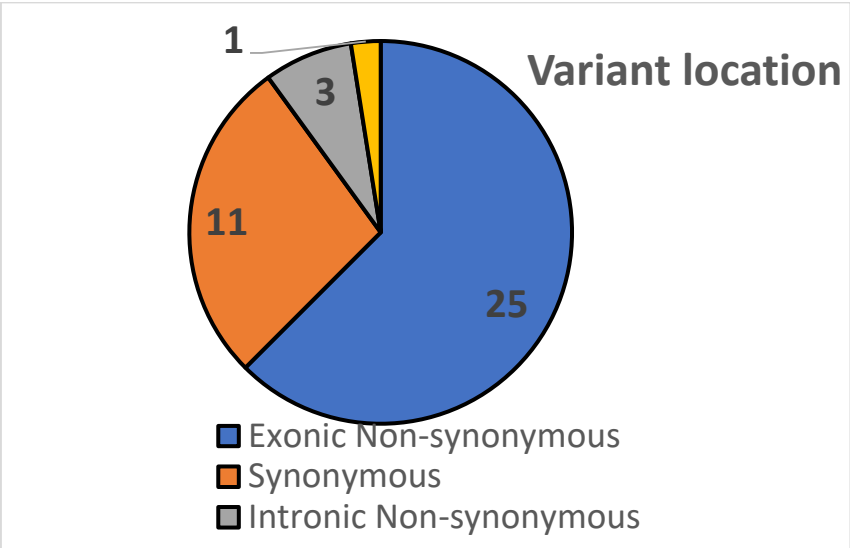



Figure. 4.18b.: Variant location identified by Targeted Panel Gene Sequencing of *BRCA1*, *BRCA2*, *CHEK2* and *PALB2*.

### *4.19. Inter population differences in the allele frequency*

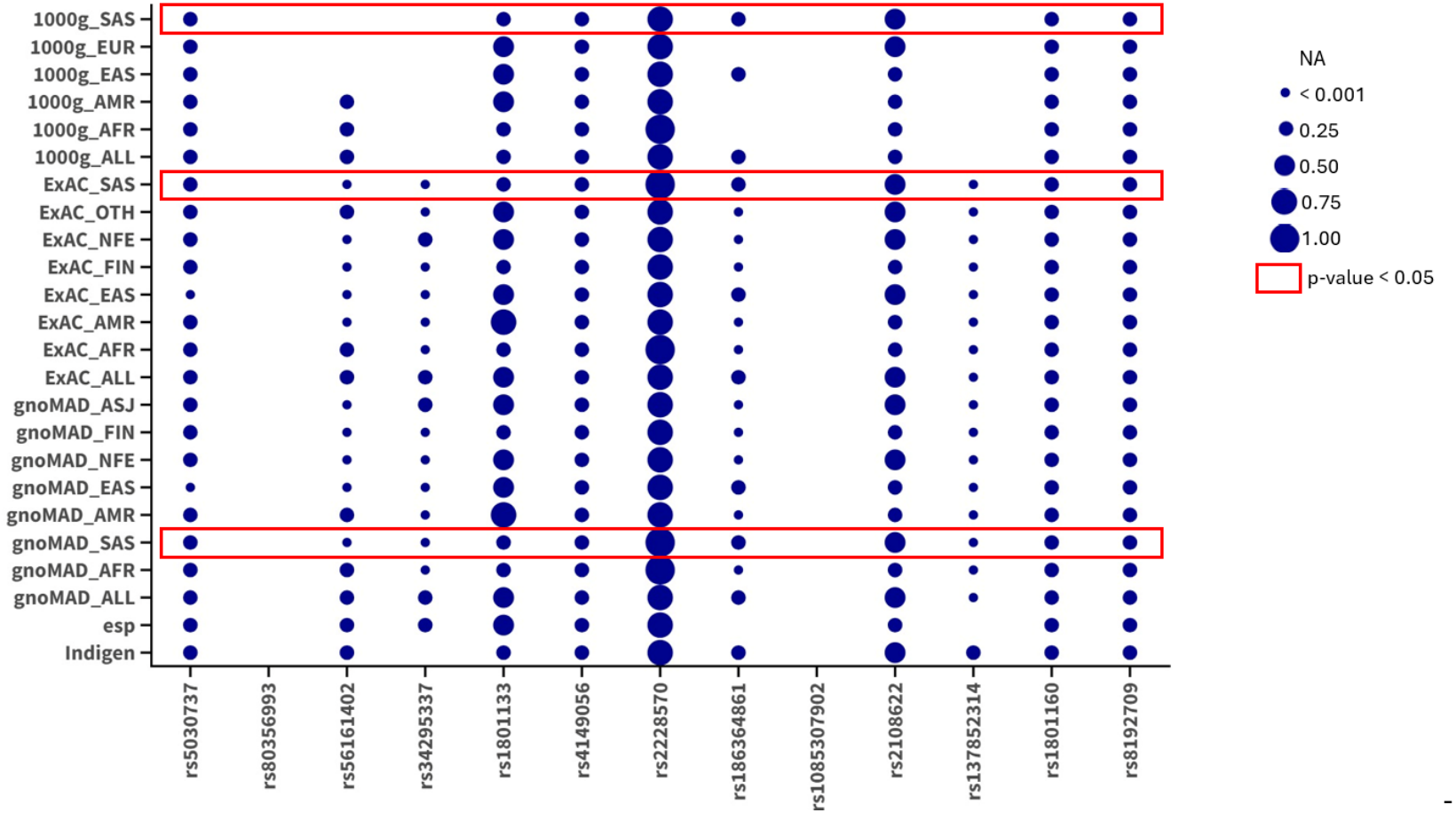



Figure.4.19.: Inter population differences in the allele frequency of important pharmacogenomic variants.

### *4.20. Cataloguing of common polymorphism*

The WES data of 17 TNBC patients were clubbed together and variants that has allele frequency lower than 0.5 as well as duplicates were excluded. Variants with no disease association in ClinVar were retained. Variants that had predicted to be benign in all three pathogenic score such as SIFT, Polyphen2 and Mutation Taster were then filtered out. These variants were then matched with healthy controls and variants that were also present in healthy controls were considered to be common polymorphic among Mizo population. A total of 35756 variants were obtained.

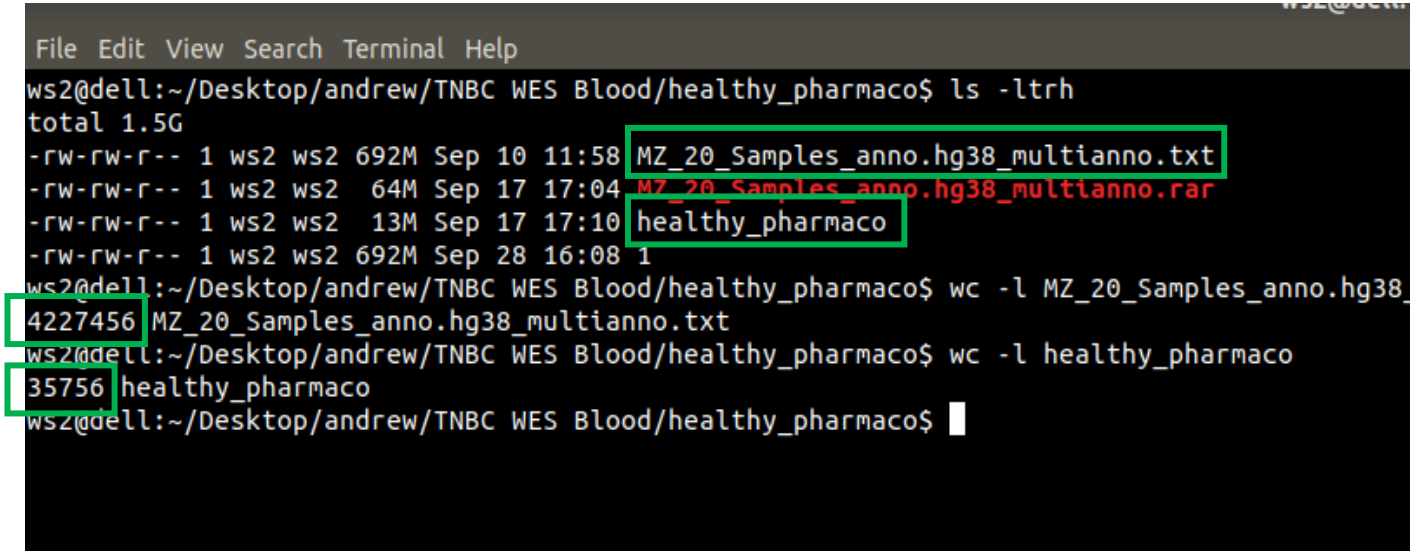



Figure. 4.20.: Cataloguing of common polymorphism with no clinical significance.

### *4.21. Principal Component Analysis*

Principal Component Analysis was performed Using R-package and maximum variants were captured in first three principal components. The populations were color-coded (African – yellow, Caucasian – green, Hispanic – light grey, East Asian – blue, South Asian – black, Mizoram - red).

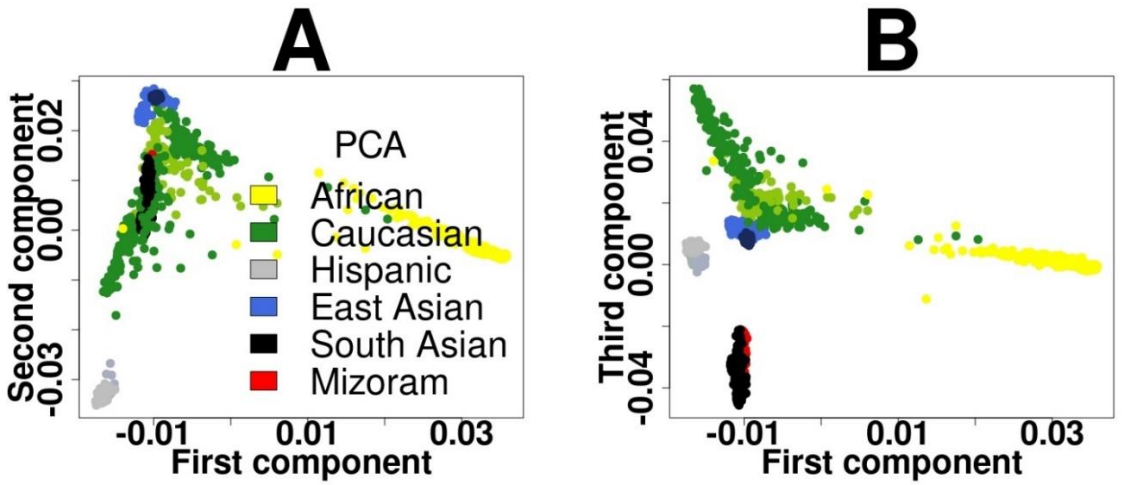



Fig. 4.21.: PCA of Mizo population with 1000 genomes. (A): Mizoram samples clustered with South Asian but overlapped with most of the population except African and Hispanic when plotted with First component with second component. Fig (B): The plot between First principal component and Third principal component clearly showed Mizo population clustered with South Asian population.

# Chapter 5

# Discussion and Conclusion

The present study involves twenty-eight TNBC patients that were admitted in Mizoram State Cancer Institute, where their clinical records, ADR related questionnaire as well as their peripheral blood were collected upon their consent. DNA was isolated and WES was performed on seventeen TNBC patients along with twenty WGS of healthy volunteers. The TNBC patients recruited for the study has a median age of 52 with the oldest being 91 years old and youngest being 24 years old, providing a wider range for studying if age has an effect in the efficacy of chemotherapy drugs administered to them. Their clinical records such as histopathology, tumour grade, MBR grade, axillary lymph node involvement, chemotherapeutic treatment administered and information regarding ADR's were also collected using a well-structured questionnaire and was compared with the mutations found among the TNBC patients.

## *5.1. Statistical analysis on ADRs with harboured mutation and clinical records*

Univariate Binary Logistic Regression was performed to reveal the association between the ADR and factors like mutations, clinical records such histopathology, grade of the tumour, axillary lymph node and the combinations of chemotherapy administered. Almost three decades ago, after *BRCA1* was cloned and extensively studied (Miki *et al.,* 1994), apart from involving in different cellular processes, its main role was deduced to be in DNA damage response (Jiang *et al.,* 2015) especially in double stranded breaks performed through Homologous Recombination (HR) or by Non-Homologous End Joining (NHEJ), thereby, providing genomic stability (Li *et al.,* 2012; Kass *et al.,* 2015). BRCA1 is well establish susceptibility gene in familial BC and Ovarian cancer (OC) (King *et al.,* 2003), and after a decade of establishing its role in BC and OC, studies have also revealed that it also has implication in other cancer such as prostate and pancreatic

cancer (Robinson *et al.,* 2015; Waddell *et al.,* 2015). Several studies have reported the presence of germline mutations in *BRCA1* in Gastric Cancer (GC) (Ichikawa *et al.,* 2018; Sahasrabudhe *et al.,* 2017; Halpern *et al.,* 2020).

The intestinal type of GC can be attributed to the Correa cascade where, prolonged gastric inflammation or gastritis leads to the cascading of evolution from metaplasia, dysplasia and finally to carcinoma (Correa *et al.,* 1992). It was also reported that there is a loss of somatic *BRCA1* allele indicating an inheritance of mutant *BRCA1* suggesting predisposition in gastric malignancy (Moiseyenko *et al.,* 2013). Studies have shown that in GC patients older than 50 years, a significant association between *BRCA1* mutation carriers and gastritis (Spearman's coefficient = 0.337, $p < 0.001$) as compared to younger GC patients was found (Avanesyan *et al.,* 2020). Similarly, it was also observed that among first degree relatives of *BRCA1* mutation carriers increases 4-fold in the risk of developing GC. This could suggest the role of *BRCA1* in developing GC, however, its mechanism of *BRCA1* related pathogenesis of GC is still undetermined and merits further investigation (Buckley *et al.,* 2022). In line with previous findings, we found that patients having *BRCA1* p.C1650R mutations are likely to have Gastritis with an OR of 22 (p-value = 0.031, 95% CI = 1.334 - 362.916). It is noteworthy that as the sample size is less, even though the p-value is significant, there is a wide range in the confidence interval.

Similarly, patients with *BRCA1*.C1650R mutations are also likely to have Leukopenia with an OR of 13.06 (p-value = 0.034, 95% CI = 1.225 - 151.045). Out of the twenty-eight TNBC patients collected for ADR information, nine patients (32.14%) were known to have leukopenia, where 2 patients (22.22%) were in mild conditions, 6 (66.66%) in moderate and 1 (11.11%) in severe condition. A retrospective matched cohort study conducted from different cancer registries within the United States of America also concludes that decrease in haemoglobin count was the most common hematologic toxicity observed and that among *BRCA1* carriers, leukopenia was the most severe ADR observed (West *et al.,* 2019). In line with this study, patients with germline *BRCA1* gene mutations were predicted to have severe ADR when treated with cyclophosphamide, taxanes and carboplatin-based therapy (Furlanetto *et al.,* 2021). Our study population did not received carboplatin, however,

they received anthracycline and cyclophosphamide-based therapy in a sequential manner.

Our analysis also revealed that patients who had Axillary Lymph Node involvement are likely to have Fatigue as an ADR with an OR of 8.889 (p-value = 0.023, 95% CI = 1.344 - 58.796). It was reported that the main and first observable side effects of neo-adjuvant chemotherapy was fatigue (Kumar *et al.*, 2019). This could mean that as our patients receives anthracycline, cyclophosphamide and taxane-based therapy that could lead to haematologic toxicity. As eight out of twenty-eight patients were tested positive for axillary lymph node involvement, where axillary lymph node involvement proved to be an important prognostic factor in TNBC patients. However, BC guidelines of Saint Gallen 2015 recommended that axillary lymph node involvement with lesser than grade 3 does not necessarily requires chemotherapeutic treatment. Study was conducted to determine the need for chemotherapy based only on axillary lymph node involvement status or in corelation with clinicopathological factors that dictates the need for chemotherapy treatment. They predicted that the patients age, size and grade of tumour, subtype and axillary lymph node involvement were all to be considered (Houvenaeghel *et al.*, 2016). This suggested that although axillary lymph node involvement is a prognostic factor, it is also crucial that we also took the age, size and grade of tumour, subtype to predict the disease-free survival rate as well as overall survival rate within the population.

### *5.2. List of total variants observed (after variant prioritization) from TNBC patients*

After performing annotation for all the seventeen TNBC WES data, filtering steps were performed accordingly as mention in the methodology section. We observed approximately 30,000 variants from each of the 17 TNBC patients, however, after hard filtering steps we identified gene variants a total of 125 variants were obtained (Table. 4.7.) and were separated according to the variants that were present in: PharmGKB database, ClinVar, Variants implicated in other disease, Variants with no rsID, and Others. A simple Venn diagram was plotted to check

which variants were present in to the following categories as mentioned earlier (Figure. 4.8.). These 125 variants were further taken for downstream analysis.

The smaller number of variants obtained is due to the fact that lesser genes were considered for pharmacogenomic analysis. This is due to the fact that there are lesser number of dedicated pharmacogenomic databases that catalogue variants with their contributions to the disease phenotype and their evidence level. There are other databases which also catalogues variants that are of pharmacogenomic importance (to name a few, apart from the database that we used): PGMD (Kaplun *et al*., 2016), Nomenclature for N-acetyltransferases (Vatsis *et al*., 1995) and Pharmacogenetics and Database of Pharmacogenomic Information in Ethnic Minority Populations (Li *et al*., 2020). Among these, the most curated and updated database is the PharmGKB database and DrugBank. Therefore, although we observed a huge amount of variant data from each TNBC patients, the filtered variants that passed the criteria as mentioned in the methodology section were less.

Among the overlapping variants, there are a total of twenty-four variants that had no rsID (i.e., not reported in dbSNP) and twenty-six variants that were reported in ClinVar with no clinical significance as well as its pathogenicity in disease conditions. The twenty-four variants that had no rsID could be a population specific variant that merits further investigations as they might or might not play a role towards disease manifestations, diagnostic and prognostic marker as they were found only in TNBC patients. *MTHFR* (rs1801133) and *SLCO1B1* (rs4149056) were present both in PharmGKB database and in ClinVar. *DPYD* (rs1801160) was also present in PharmGKB database but not in ClinVar.

### *5.3. Variants that were present in PharmGKB database*

From the WES data analysis of TNBC patients, that met the analytical criteria with exhaustive filtering step, we observed four variants that were also found in PharmGKB database namely:

### *5.3a. MTHFR*

Methylenetetrahydrofolate reductase (MTHFR) is a key gene that is responsible for folate metabolism that converts 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate which is the primary form of folate that exist as a circulatory water-soluble vitamin (Vitamin B9) (Bethke *et al.,* 2008). 5,10-methylenetetrahydrofolate is utilized by thymidylate synthase and deoxyuridine monophosphate (dUMP) undergoes methylation producing deoxythymidine monophosphate (dTMP) which is a source of thymidine in DNA synthesis and DNA repair processes and is also required to remethylate homocysteine to methionine. Mutation in MTHFR can lead to loss of function or lower rate of conversion of 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate thus accumulation of 5,10-methylenetetrahydrofolate occurs (Hasan *et al.,* 2019). Insufficient folate levels or diminished enzyme activity within the folate metabolism pathway can lead to a shortage of the substrate required for methionine synthase, thereby impacting the re-methylation process and resulting in elevated homocysteine levels in the bloodstream. Heightened concentrations of homocysteine in plasma have been associated with various forms of human cancer such as BC (Bravatà *et al*., 2014), colorectal (Matsuo *et al*., 2005), prostrate (Singal *et al*., 2004) and leukemia (Robien *et al*., 2003).

The most common MTHFR gene mutation that has impact on breast cancer are p.C677T and the p.A1298C variants. p.C677T variant and its genotype CT and TT was found to increase risk of BC (Tsang *et al.,* 2015) in different non-Asian populations like Jordanian (Awwad *et al.,* 2015), Iranian (Hesari *et al.,* 2019) and Latin American (Meneses-Sanchez *et al.,* 2019) and also in mixed population including Asian and Caucasian (He *et al.,* 2017; Li *et al.,* 2014; Liu *et al.,* 2016; Ajaz

*et al.*, 2021). However, there are also studies that found no correlation between CT and TT alleles with increased risk of developing BC among the Brazilian (Durán *et al.*, 2021) and North American population (Houghton *et al.*, 2019). Moreover, study conducted in Indian population did not find any association indicating these genotypes increase the risk of BC (Pooja *et al*., 2015). These contradicting findings from different population indicates that there could be a population specific mutational signatures that differentiates them with respect to MTHFR gene mutations and that identification of such ethnic specific mutation is of an utmost importance.

Our study identified a variant (G>A, c.C788T, p.A222V) in the MTHFR gene that confers toxicity towards capecitabine and fluorouracil with level 4 evidence in four TNBC patients (23.52%). PharmGKB database had reported that genotype AA has the highest risk of toxicity followed by GA genotype. This indicates there is increase risk of toxicity when treated with capecitabine and fluorouracil than the wild type genotype. All TNBC patients in our study groups received 5- fluorouracil which is a thymidylate synthase inhibitor. This MTHFR variant might confer loss of function or reduced enzyme activity in converting 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate thereby increase in the concentration of 5,10-methylenetetrahydrofolate which is a precursor molecule in the subsequent production of homocysteine levels in the bloodstream (Hasan *et al*., 2019). It is established that high concentrations of homocysteine in plasma have been associated with various forms of human cancer such as BC (Bravatà *et al*., 2015). However, it is noteworthy that the functional implication of this variant is not known and further experiments are required to establish its function towards toxicity against capecitabine and fluorouracil. Similarly, germline variant of MTHFR p.A222V was predicted to be deleterious in any two of the three predicting tools (SIFT, PolyPhen2 and Mutation taster) and was also reported in PharmGKB database and ClinVar that confers toxicity. Nonetheless, the allele frequency of the genotype GA has yet to be studied with a larger number of patients to understand whether Mizo population has been predisposed to this variant.

### *5.3b. SLCO1B1*

The Solute Carrier (SLC) superfamily of transporters includes hundreds of family members and are localized in the membranes of different organs, involving in absorption, distribution and elimination of vital nutrients, toxins and therapeutic drugs (Roth *et al.,* 2012). Organic anion-transporting polypeptides encoded by SLC family has a well-established substrate from many of the important clinical drugs. Previous research has indicated correlations between genetic variations in SLCO genes and the pharmacokinetics of their substrates (Joyce et al., 2015). One prevalent SNP (rs4149056; T>C, p.V174A) in SLCO1B1 gene has been associated with reduced transporting efficacy of OATP1B1. Consequently, this SNP leads to significantly elevated plasma levels of numerous statins. Similarly, they also found that specific variants of the same gene were linked to accelerated methotrexate clearance, thus increasing the chance of gastrointestinal toxicity during methotrexate therapy in pediatric acute lymphoblastic leukemia (Niemi *et al.,* 2011).

In this study, none of the patients received Methotrexate or statin thus we cannot find any relationship between SLCO1B1 variant (rs4149056; T>C, p.V174A). However, the same variant was studied for younger female with BC (age<45) to understand the variant involvement and developing of chemotherapy-induced amenorrhea in the patients. They found that this variant is a predictor of developing chemotherapy-induced amenorrhea (Reimer *et al.,* 2016). Nonetheless, PharmGKB database has catalogued to have a toxic effect when BC patients are treated with cyclophosphamide with a level of evidence 3.

### *5.3c. DPYD*

Within the cellular environment, 5-fluorouracil (5-FU) undergoes a series of enzymatic reactions, giving rise to multiple active metabolites. One such metabolite is fluorodeoxyuridine monophosphate, which acts as a competitive inhibitor of thymidylate synthase. This inhibition disrupts the conversion of deoxyuridine monophosphate to deoxythymidine monophosphate, a pivotal precursor of thymine (Shahrokni *et al.,* 2009; Scartozzi *et al.,* 2011). Consequently, an imbalance in

nucleotide availability crucial for the proliferation of cancer cells occurs (Petak *et al.*, 2000), leading to aberrant cell growth and finally apoptosis (Anan *et al*., 2003). Moreover, 5-FU (also a pyrimidine analog) has the capability to integrate into DNA, instigating DNA damage repair mechanisms and initiating apoptotic pathways (Shahrokni *et al.,* 2009). *DPYD* encodes Dihydropyrimidine dehydrogenase (DPD) which is responsible in the catabolic process leading to the clearance of 5-FU and is responsible for 5-FU toxicity within the cells.

Previous studies had identified four DPYD variants which are proven to be clinically related with 5-Fu toxicity: rs3918290, rs56038477, rs67376798 and rs55886062 (van Kuilenburg *et al*., 2002; Amstutz *et al.,* 2017; Vreken *et* al., 1996; Henricks *et al.,* 2015,). Recently, the largest comprehensive study of DPYD variants among South Asian population was conducted with respect to 5-Fu toxicity and found that fifteen variants were of importance and merits further clinical investigations (Hariprakash et al., 2017). The variant (p.V732I) identified in our study (rs1801160) was also identified in this study with the highest allele frequency among all the studied cohorts but was not so as compared to non-Asian cohort, meaning this variant is more prevalent among the Asians. It was also observed that there was no association between the variant and the toxic effects from 5-Fu as there was no significant changes in DPD deficiency which was also previously reported by other study group from Chinese population (He *et al.,* 2008). Moreover, PharmGKB database also reported that rs1801160 do not have any association with 5-Fu toxicity and that although, there is an amino acid substitution (p.V732I), it was assigned to have normal function by Clinical Pharmacogenetics Implementation Consortium (CPIC).

### 5.3d. *CPY2B6*

The association between steroid hormones such as estradiol and testosterone with breast cancer risk is complex. Testosterone has been observed to potentially exert a greater impact on breast cancer risk in postmenopausal women compared to estradiol (Key *et al.,* 2003). This influence could stem from testosterone's dual role:

it can either act as a precursor to estrogen within the breast by undergoing conversion, or directly stimulate the proliferation of breast cells (Poutanen *et al*., 1995).

The enzyme cytochrome P450 2B6 (CYP2B6) plays a critical role in metabolism of testosterone. It facilitates the 16-alpha- and 16-beta-hydroxylation processes, which are pivotal steps in testosterone deactivation. This enzymatic activity is significant as it modulates the balance of testosterone and its metabolites in the body, thereby impacting breast cancer risk (Imaoka *et al.,* 1996). Understanding the involvement of enzymes like CYP2B6 in steroid hormone metabolism provides insights into the underlying mechanisms of hormone-related diseases such as breast cancer. Certain CYPs (the cytochrome P450 family) have also been implicated in TNBC, particularly those involved in activating fatty acids (Apaya *et al.,* 2016), inactivating calcitriol (Flanagan *et al*., 2003), and inactivating retinoic acid (Helvig *et al*., 2011). CYP2B6 is also involved in the metabolisim of some xenobiotics, such as the anti-cancer drugs cyclophosphamide and ifosphamide (Mangó *et al*., 2023). CYP2B6 variant (c.516G and c.85G) referred as CYP2B6*6, was found to decrease the activity of CYP2B6 enzyme, thus, elevating the risk of BC (Justenhoven *et al*., 2014).

Our study reveals that CYP2B6 variant (c.C64T, p.R22C) was present in four TNBC patients. This was reported in PharmGKB database that confers toxicity towards cyclophosphamide with level of evidence as 3. It was reported that the variant homozygous T allele was associated with increased likelihood of dose delay when treated with cyclophosphamide and doxorubicin in women with Breast Neoplasms as compared to homozygous allele C (Bray *et al*., 2010; Kalra *et al*., 2018). However, CT allele was reported to be associated with decreased time to progression when treated with cyclophosphamide and doxorubicin, thus conferring a positive impact on treating with cyclophosphamide (Bray *et al*., 2010).

### *5.4. Variants present in ClinVar*

#### *5.4a. BRCA1*

Over two decades ago, the correlation between the *BRCA1* and *BRCA2* genes and susceptibility to breast and ovarian cancer was initially established (Miki *et al.*, 1994). Subsequently, particular germline mutations within these genes have been identified, indicating heightened risks for various other types of cancers, such as prostate, colorectal, stomach, and pancreatic cancers (Cavanagh *et al.*, 2015; Mersch *et al.*, 2015). It's been found that mutations in the *BRCA1* genes are accountable for approximately 20% of familial cases of breast and ovarian cancers (Noh *et al.*, 2015). *BRCA1* has many interacting partners to fulfil its DNA repair functions, The C-terminal end of the *BRCA1* (BRCT domain) (Clark *et* al., 2012) and the N-terminal which is frequently interact with BRCA1-associated RING domain protein 1 (BARD1) (Hashizume *et al.*, 2001) and apart from this, it can also interact with *PALB2* and BRCA2, a stable complex called as BRCA1-PALB1-BRCA2 (Zhang *et al.*, 2009). The domains of BRCA1 and its interacting partners through the BRCT domain, such as BRCA1/PALB2/BRCA2 complex, BRCA1-RAP80-Abraxas complex, BRCA1-BACH1complex, and BRCA1-CtIP complex play a crucial role in DNA repair mechanism by homologous recombination thereby maintaining the genome stability (Fu *et al.*, 2022).

Among the common germline mutations in the BRCA1 gene are 5382insC, 185delAG, 3819del5, and 4153delA (Wang et al., 2012). In the Western population, around 5% of breast cancer patients may harbour heritable mutations predisposing them to cancer, with BRCA1 mutations being the most prevalent (Tung et al., 2015). This mutation prevalence tends to be higher among Ashkenazi Jews (Zhu et al., 2016). Interestingly, the occurrence of *BRCA1/2* mutations in Asian populations is generally lower compared to that in Caucasian populations (Zhang et al., 2012). The missense mutation, c.5089T>C variant in *BRCA1* leads to the substitution of Cysteine with Arginine at amino acid position 1697 (p. Cys1697Arg). Notably, this variant is from gnomAD v3.1. Moreover, this missense variant lies within a crucial functional domain, and according to the BayesDel computational predictor (noAF), it

receives a score of 0.40, surpassing the recommended threshold of 0.28 for predicting its impact on BRCA1 function via. protein alteration. Additionally, the SpliceAI predictor score of 0.05 indicates that the variant is not involved in splicing. Furthermore, many studies reported that this variant is a pathogenic variant (Fernandes et al., 2019; Petitalot et al., 2019; Findlay et al., 2018; Parsons et al., 2019). The variant was also subjected to Saturation Genome Editing (SGE) assay and resulted to have a score of -2.15, which is classified as loss of function (Findlay et al., 2018).

This variant was absent in gnomAD v3.1, and when annotated does not have any allele frequency in population databases such as ExAc and 1000 genomes (at the time of annotation), it could possibly more predisposed towards Asian population and that provides a platform for this variant to be a diagnostic biomarker. For it to be used as a biomarker, larger number of samples will be required to prove its validity.

### *5.4b. VDR*

The function of vitamin D relies on its binding to the Vitamin D Receptor (VDR), which then facilitates various cellular processes (Spina *et al*., 2003). One such function is the potential tumour-suppressive role of the vitamin D-VDR complex, which operates by influencing the TGF-β signalling pathway (Jiang *et al*., 2013). The mechanism by which VDR activates transcription at the genomic level involves its recognition and binding to specific regions called vitamin D response elements (VDREs) located within the promoter regions of target genes (Feldman *et* al., 2014). Variations in the VDR gene have been linked to alterations in cancer risk and prognosis (Köstner *et al*., 2009). Genetic variations in VDR may impact the risk of breast cancer by altering its expression and function within breast cells. Furthermore, over expression of VDR messenger Ribonucleic Acid (mRNA) had a poor prognosis in BC (Nam *et al*., 2022).

One of the most common and well-studied variant in VDR is the "c.T2C:p.M1?", and had been also reported to have an increased risk of developing BC (McCullough *etal.,* 2007; Akilzhanova *et* al., 2014; Kazemi *et al.,*2022) and

Ovarian Cancer (Lurie *et al*., 2011; Chen *et al.,* 2018) This variant is also known as the *FokI* polymorphism due to the presence or absence of a restriction enzyme site. The Fok1 f allele found in the 5' promoter region of the VDR (rs2228570) causes the VDR protein to be elongated by three amino acids, resulting in decreased effectiveness as a transcriptional activator (Uitterlinden *et al*., 2004). As we also observed the same variant in sixteen (16) TNBC patients (88.88%), this variant may be predisposed to the Mizo population and merits further investigation with larger samples. It is noteworthy that, this variant does not present in any of the healthy population and might be an important genomic biomarker for the population. Since, VDR has different cellular roles and is involved in many diseases, investigating the interplay between vitamin D, the VDR, and cancer risk is an active area of research, with ongoing efforts to understand the underlying mechanisms and clinical implications.

### *5.5. Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1*

A total of 39 variants unique were observed from 28 TNBC patients (Figure. 4.17a). Of that 25 variants were nonsynonymous and exonic in function, 11 were synonymous and 4 was a modifier with 1 variant as 5ÚTR in function and 3 were in the intronic region (Figure. 4.17b). Out of these we found only one significant finding, which is the BRCA1 (c.5089T>C, p. Cys1697Arg) variant in leads to the substitution of Cysteine with Arginine at amino acid position 1697. This variant is already discussed above

### *5.6. Inter-population differences in the allele frequency of important pharmacogenomic variants*

Identification of these genomic changes can help us to better understand disease manifestations, progression and its way to tackle with significantly lesser side effects. According to precision medicine, these variation in the genomes could be utilized for disease treatment and prevention and considers individual variability

in genes, environment, and lifestyle for each person (Terry, 2015). Breast cancer is one of the most studied cancers globally and the biomarkers for early detection and its precise treatment had also been developed (Mehrgou and Akouchekian, 2016). The treatment advised and its disease-free survival and its relapse can also be attributed to its different subtypes (Carol and Vincent, 2014; Onitilo *et al*., 2009). Although, extensive studies had been conducted in the molecular mechanisms, diagnosis and prognosis for breast cancer leading to a fruitful clinical outcome, one of the subtypes of breast cancer presents a challenging scientific exploration to have a better early diagnosis, and its treatment regime (Collignon *et al.*, 2016; Zaharia).

Upon comparing the allele frequency of important pharmacogenomic variants using Fisher exact test we found that there is significant association between the rsIDs selected with that of the samples present in ExAc and gnomAD databases. ExAc consists of only exome sequencing data and was previously a stand-alone database, and later was merged to gnomAD. ExAC has a catalogue of approximately 7.4 million exonic variants with no report on structural variation, while gnomAD consists of 241 million genomic variants along with 333,470 structural variants. Thus, it is showing significant association with these two databases (currently are now merged) (Figure. 4.18.). However, it is also evident that these selected important pharmacogenomic variants differ in the allele frequency across other databases, indicating a mixed allele frequency distribution. This suggest that the variants observed in our study could also mean that it might be a population specific which merits further investigation with larger number of samples.

### *5.7. Pathway analysis for Druggable target*

From the WES data (annotated file), all the genes were clubbed from each of the seventeen TNBC patients and the duplicated genes were removed. This list of genes was used for pathway analysis using DAVID with the latest DAVID Knowledgebase v2023q2 (at the time of analysis). KEGG pathway analysis revels top 14 pathways that were implicated with the genes analysed; including ABC transporters, Chemical carcinogens – DNA adducts, Chemical carcinogenesis –

receptor activation, mineral absorption, MAPK signalling pathway. Similarly, in Reactome, the analysis showed 45 possible pathways: including Drug Absorption Distribution Metabolism and Excretion (ADME), Transport of small molecules, Xenobiotics, CYP2F1 reactions, ABC-family proteins mediated transport, Metabolic disorder of biological oxidation enzymes.

Most of the pathway observed are related to cancer, that is, genes responsible for inducing carcinogenesis. Upon identifying the Drug-Interaction Function and analysed using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions. A total of 11 druggable categories and 11 druggable genes were identified. These druggable categories includes Tumour suppressor, DNA repair mechanism, Transporter, Cell surface receptors etc. This suggest that the variants found in our study has an important role in being a target for developing new or utilizing a already existing chemotherapeutic drugs. MTHFR (rs1801133), SLCO1B1 (rs4149056) and CYP2B2 (rs8192709) could act as a prognostic marker while, BRCA1 (rs80356993) and VDR (rs2228570) could potentially acts as predictive biomarker for TNBC within the population.

### *5.8. Cataloguing of common polymorphic variants among Mizo population*

NGS data encompasses a huge information that covers approximately 60 million base pairs which is about 2% of the genome that includes variants that are of significant value in disease association as well as variants that do not play a role in diseases. Usually, NGS experimental design with a coverage 100x and 150 pair-end reads generates approximately 10Gb of data per sample run which can be translated to about 30,000 to 40,000 variants per sample. So, there is a huge challenge in traversing such a huge date to pick out important and significant variants that might play a key role in the disease studied. Therefore, there are a lot of bioinformatic tools available to eliminate those variants that might be a simple polymorphism for the population. However, there are variants which is more prone to a specific population and such variants could also be a predisposed variant that might confer risk of developing certain disease, toxicity towards drugs administered or even a causal

factor for the population. The concept of predisposition has a greater effect on certain populations that practises endogamy and mizo population are a great example of populations that had practise endogamy since time immemorial. Thus, for a population such as mizo, there is a need for identifying variants that are common polymorphic in nature and does not contribute to disease association.

These identified variations hardly bring out the true nature of the diversity that exists within the Indian population. Nevertheless, there are also few Mizo genetic data in terms of the exonic region and in mitochondrial DNA that are produced inhouse (Freda *et al.*, 2020; Andrew *et al.*, 2020; Vanlalhruaii *et al.*, 2020). This suggested that there is a difficulty in mapping, calling of variants, annotation and its interpretation of variants to its functional and clinical impacts. It is important to know the genomic background of a population so as to provide a targeted treatment specific to their own requirements. Such goal can now be achieved by the advancement of Next generation Sequencing methods. This provides a platform where the technology identifies the DNA alterations that also means that it also identifies the gene where the variants are present. There could be a difference in the genomic background in different populations, thus identifying such differences is of the urgent need.

### *5.9. PCA analysis*

In general, maximum variants were captured in first three principal components. Using R with principal component 1 and 2 plotted graph, the populations were color-coded (African – yellow, Caucasian – green, Hispanic – light grey, East Asian – blue, South Asian – black, Mizoram - red). Mizoram samples clustered with South Asian which was evident in Fig (A). Fig (B) was the plot between principal component 1 and principal component 3 which showed Mizoram population clustered with South Asian clusters. Although Mizo population was clustered with South Asian, it overlapped with the Hispanic population too. This could be because of the two components used were the first and the second, which tallies majority of the variants used, and also it could be because of the lesser sample

size as compared to the other population. For this PCA plot, the first component used 90 – 95% of the variants, second component used 85 – 90% and the third component used 75 – 85% from the total variants observed from the called variants.

The earliest genetic diversity study of Mizo population we could obtain was a comparison between Mongoloid, Sino-Indian speaking tribal groups (Toto, Mizo, Tharu), and a morphologically proto-Australoid, Austro-Asiatic speaking tribal group (Ho) as and out group. This study uses polymorphisms at 25 loci of which 8 were insertion/deletion poly-morphisms and the remaining 17 were RFLPs where they identified Mizo were not related to any of the other mongoloid tribal groups (Chakrabarti *et al*., 2002) and suggested that morphologically similar mongoloid might originated or migrated separately. Recently (while writing this manuscript), a Genome wide data comprising of 100 Mizo samples were collected and the genetic diversity was studied. They found that Mizo population were although related to East and South Asians, but form a distinct cluster (Bankura *et al*., 2024). The difference in clustering of our study and their study could be attributed to the number of samples used in each study. The limitations of our study is that we did not choose specific tribes, but we have to select TNBC patients and twenty healthy individuals.

## Chapter 6

## Summary

1. A total of 28 TNBC patients who are admitted in MSCI were recruited for the study after obtaining a signed consent form.
2. The TNBC patients recruited for the study has a median age of 52 years with the oldest being 91 years and youngest being 24 years providing a wider range for studying if age has an effect in the efficacy of chemotherapy drugs administered
3. Their clinical records such as histopathology, tumour grade, Modified Bloom-Richardson (MBR) grade, axillary lymph node involvement, chemotherapeutic treatment administered was collected.
4. 20 individuals who are self-proclaimed healthy were taken as healthy controls with the possibility of using their genomic data to study the diversity with other populations.
5. These 20 individuals were carefully selected based on: genetically unrelated, different sub-tribes representing the different districts in-order to maintain sample diversity.
6. This study aims to corelate between the genomic background of TNBC patients with the manifestations of different ADRs and their clinical records.
7. To study the possible ADRs with different treatment regime, a well-structured questionnaire was formulated to follow the uniform Clinical Practice Guidelines (CPG) and the National Comprehensive Cancer Network (NCCN) was used.
8. The structured questionnaire contains information such as personal details, date of diagnosis, histopathology, tumour location and grade, MBR grade, auxiliary lymph node involvement, types of chemotherapy drugs administered and no of cycles, TNM staging and different types of ADRs.
9. The questionnaire was followed-up in 6 months interval to check their vital status and to record any complications they had endured.
10. In order to understand the relation of genomic variants, clinical and histopathological records and ADR, Spearman rank correlation co-efficient was used

for significant level (p-value) <0.05 and confidence level of 95% were maintained using IBM Statistical Package for Social Sciences (SPSS) version 22.0.

11. WES was performed on17 TNBC samples and Agilent SureSelectXT target enrichment System was used for library preparation, and was sequenced in Illumina Hi-Seq2500 at 100X depth.

12. After performing quality control, the sequencing reads were aligned to the reference genome (GRCh38), variant calling was done using GATK Haplotype caller and ANNOVAR was used for annotating the variants.

13. WGS was performed on 20 healthy individuals using TruSeq DNA PCR free library preparation kit and sequenced on Illumina NovaSeq 6000.

14. After performing quality control, the sequencing reads were aligned to the reference genome (GRCh38), Sentieon was used for joint variant calling and ANNOVAR was used for annotating the variants.

15. The Very Important Pharmacogenes" was extracted from PharmGKB database. The list contains 69 genes.

16. Genes responsible for transport and metabolizing of 5-Fluorocil, Epirubicin, Cyclophosphamide and Paclitaxel with Capecitabine, Doxorubicin, Docetaxel were extracted from DrugBank and through literature search were identified and listed and named as metabolizing genes.

17. We matched the annotated file of TNBC patients with PharmGKB gene list and the same with metabolizing gene list, and we retained only those variants which are exonic and splice site in function, then we select variants that are predicted to be pathogenic/deleterious by any two of the pathogenicity predicting tools.

18. A total of 111 genes were selected for pathway analysis using KEGG and Reactome in DAVID.

19. The Maftools in R-package was used to find the Drug-Interaction Function using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions.

20. The allele frequency of the important pharmacogenomic variants we observed were retrieved from IndiGenomes database and was grouped into five groups as 0.0, 0.25, 0.50, 0.75 and 1.0, after which Fisher Exact test was performed on the allele frequencies to find the corelation with other population using databases such as The

1000 Genomes Project (1000g), The Exome Aggregation Consortium (ExAC), The Genome Aggregation Database (gnomAD) and Exome Sequencing Project (esp6000).

21. The WES data of 17 TNBC patients were clubbed together and variants that has allele frequency lower than 0.5 were excluded. ClinVar database was used to check whether these filtered variants have any association with any disease. Variants that had predicted to be benign in all three pathogenic score such as SIFT, Polyphen2 and Mutation Taster were then filtered out. These variants were then matched with healthy controls and those that were also present in healthy controls were considered to be common polymorphic among mizo population.

22. Principal component analysis (PCA) was used to identify the genetic and geographical distances between population by preserving their covariance matrixes. 1000Gphase1.snps vcf and their indexed files were used.

23. Linkage disequilibrium (LD) pruning of our sample and 1000 genomes was performed using reference genome GRCh38, and the common variants between Mizoram population and 1000 genome data were extracted and merged using bcftools. Eigen vectors were computed using Plink software.

24. A total of 30 ADRs were observed from 28 TNBC patients.

25. The top 5 ADR observed are Alopecia (89.28%), Nausea (75%), Drowsiness (64.28), Change in appetite (53.57%) and Constipation (46.42%)

26. The most severe ADR was Alopecia and majority of the of ADR severity observed was moderate.

27. Analysis showed that patients having BRCA1.C1650R mutations are likely to have Gastritis with an OR of 22 (p-value = 0.031, 95% CI = 1.334 - 362.916).

28. Patients with CYP4F2 mutations are also likely to Leukopenia with an OR of 13.06 (p-value = 0.034, 95% CI = 1.225 - 151.045). On the other hand, patients with the same mutation are less likely to have Drowsiness as an ADR with an OR of 0.074 (p-value = 0.034, 95% CI = 0.007 - 0.817).

29. Univariate Binary Logistic Regression was performed to reveal the association between the ADR and factors like mutations, clinical records such histopathology, grade of the tumour, axillary lymph node and the combinations of chemotherapy administered

30. Our analysis also revealed that patients who had Axillary Lymph Node involvement are likely to have Fatigue as an ADR with an OR of 8.889 (p-value = 0.023, 95% CI = 1.344 - 58.796).

31. After hard filtering steps WES analysis reveals a total of 125 variants were observed that could potentially play a role in clinical settings using PharmGKB database and DrugBank.

32. We identify 4 variants that were having toxic effects with chemotherapeutic drugs and were present in PharmGKB database.

33. DPYD (p.V732I, rs1801160) showed toxic effect to patients that were treated with fluorouracil with level of evidence score as 1A.

34. MTHFR (p.A222V, rs1801133) showed toxic effect to patients that were treated with fluorouracil and capecitabine with level of evidence score as 4.

35. SLCO1B1 (p.V174A, rs4149056) showed toxic effect to patients that were treated with cyclophosphamide; docetaxel; doxorubicin; epirubicin; fluorouracil with level of evidence score as 3.

36. CYP2B6 (p.R22C, rs8192709) showed toxic effect to patients that were treated with cyclophosphamide with level of evidence score as 3.

37. We also observed 11variants that were showing pathogenic in three predicting tool and are also reported in ClinVar

38. rs5030737 (MBL2:p.R52C), rs80356993 (BRCA1:p.C1650R), rs56161402 (TPMT: p.R200H), rs56161402 (TPMT: p.R200H), rs34295337 (ERCC3:p.R109X), rs1801133 (MTHFR: p.A263V), rs4149056 (SLCO1B1: p.V174A), rs2228570 (VDR: p.M1?), rs186364861 (NUDT: p.V18I), rs1085307902 (BRCA1: p.Q79X), rs2108622 (CYP4F2: p.V433M), rs137852314 (G6PD: p.G163S ) were found in ClinVar.

39. rs2228570 (VDR: p.M1?) was the most prevalent variant found in 16 samples, followed by rs2108622 (CYP4F2: p.V433M) found in 8 samples.

40. rs80356993 (BRCA1:p.C1650R) was reported in many other studies and could be a diagnostic biomarker for the population as 3 patients harbours this mutation. However, validation by larger sample size will be required

41. Among the overlapping variants, there are a total of twenty-four variants that had no rsID (i.e., not reported in dbSNP) and twenty-six variants that were reported

in ClinVar with no clinical significance as well as its pathogenicity in disease conditions.

42. Using the observed important pharmacogenomic variants obtained in our study, we compare the allele frequency with different databases such as 1000 genomes, ExAc, ESP and gnomAD and found no association leading to a notion that it could be a population specific variant.

43. KEGG pathway analysis revels top 14 pathways that were implicated with the genes analysed; including ABC transporters, Chemical carcinogens – DNA adducts, Chemical carcinogenesis – receptor activation, mineral absorption, MAPK signalling pathway.

44. Reactome pathway analysis showed 45 possible pathways: including Drug Absorption Distribution Metabolism and Excretion (ADME), Transport of small molecules, Xenobiotics, CYP2F1 reactions, ABC-family proteins mediated transport, Metabolic disorder of biological oxidation enzymes.

45. Drug-Interaction Function was identified using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions. A total of 11 druggable categories and 11 druggable genes were identified.

46. Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1reveals 39 variants with their frequency of prevalence among TNBC patients. Out of which 25 non-synonymous variants were in exonic region, 11 were synonymous, 3 non-synonymous in intronic region and 1 in 5'- UTR region.

47. The WES data of 17 TNBC patients were clubbed together and variants that has allele frequency lower than 0.5 as well as duplicates were excluded. Variants with no disease association in ClinVar were retained. Variants that had predicted to be benign in all three pathogenic score such as SIFT, Polyphen2 and Mutation Taster were then filtered out. These variants were then matched with healthy controls and variants that were also present in healthy controls were considered to be common polymorphic among Mizo population. A total of 35756 variants were obtained.

48. Principal Component Analysis was performed Using R-package and reveals that using First principal component and Third principal component, Mizo population clustered with South Asian population.

**BIBLOGRAPHY**


1. Hatton IA, Galbraith ED, Merleau NSC, Miettinen TP, Smith BM, Shander JA. The human cell count and size distribution. Proc Natl Acad Sci U S A. 2023 Sep 6;120(39).
2. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. Cell, 100(1), 57-70.
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394-424.
4. Sung H, Ferlay J, Siegel RL, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.
5. Henry NL, Shah PD, Haider I, Freer PE, Jagsi R, Sabel MS. Chapter 88: Cancer of the Breast. In: Niederhuber JE, Armitage JO, Doroshow JH, Kastan MB, Tepper JE, eds. Abeloff's Clinical Oncology. 6th ed. Philadelphia, Pa: Elsevier; 2020.
6. American Cancer Society. Accessed at https://www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer.htmlt on 10th October 2023.
7. Sharma GN, Dave R, Sanadya J, Sharma P, Sharma KK. Various types and management of breast cancer: an overview. J Adv Pharm Technol Res. 2010 Apr;1(2):109-26.
8. Perou CM, Sorlie T, Eisen MB, *et al*: Molecular portraits of human breast tumours. Nature 406:747-752, 2000.
9. Sotiriou C, Neo SY, McShane LM, *et al.* 2003. Breast cancer classification and prognosis based ongene expression profiles from a population-based study. Proc Natl Acad Sc i U S A 100:10393-10398.
10. Sørlie T, Perou CM, Tibshirani R, *et al*., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sc i U S A 98:10869-10874.

11. Sorlie T, Tibshirani R, Parker J, *et al*., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-8423.

12. Zhao H, Langerod A, Ji Y, *et al*., 2004. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 15:2523-2536.

13. Calza S, Hall P, Auer G, *et al*. 2006. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 8: R34.

14. Hu Z, Fan C, Oh DS, *et al*. 2006. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96.

15. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol. 2008 May 20;26(15):2568-81.

16. Bergamaschi A, Kim YH, Wang P, et al: Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45:1033-1040, 2006

17. Chin K, DeVries S, Fridlyand J, et al: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10:529-541, 2006

18. Vincent-Salomon A, Gruel N, Lucchesi C, et al: Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 9:R24, 2007

19. Turner NC, Reis-Filho JS: Basal-like breast cancer and the BRCA1 phenotype. Oncogene 25:5846-5853, 2006

20. Reis-Filho JS, Pinheiro C, Lambros MB, et al: EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. J Pathol 209:445-453, 2006.

21. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006 Jun 7;295(21):2492-502.

22. Lehmann, B.D.; Bauer, J.A.; Chen, X.; Sanders, M.E.; Chakravarthy, A.B.; Shyr, Y.; Pietenpol, J.A. Identification of human triple-negative breast cancer

subtypes and preclinical models for selection of targeted therapies. J. Clin. Investig. 2011, 121,2750–2767.

23. Lehmann, B.D.; Jovanovic, B.; Chen, X.; Estrada, M.V.; Johnson, K.N.; Shyr, Y.; Moses, H.L.; Sanders, M.E.; Pietenpol, J.A. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS ONE 2016, 11.

24. Burstein, M.D.; Tsimelzon, A.; Poage, G.M.; Covington, K.R.; Contreras, A.; Fuqua, S.A.; Savage, M.I.; Osborne, C.K.; Hilsenbeck, S.G.; Chang, J.C.; et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin. Cancer Res. 2015, 21, 1688–1698.

25. Jiang, Y.Z.; Liu, Y.; Xiao, Y.; Hu, X.; Jiang, L.; Zuo, W.J.; Ma, D.; Ding, J.; Zhu, X.; Zou, J.; et al. Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: The FUTURE trial. Cell Res. 2021, 31,178–186.

26. Stover, D.G.; Parsons, H.A.; Ha, G.; Freeman, S.S.; Barry, W.T.; Guo, H.; Choudhury, A.D.; Gydush, G.; Reed, S.C.; Rhoades, J.; et al. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. J. Clin. Oncol. 2018, 36, 543–553. [CrossRef] [PubMed]

27. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012, 486, 395–399.

28. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. Nature 2012, 490, 61–70.

29. Sharma P, Klemp JR, Kimler BF, Mahnken JD, Geier LJ, Khan QJ, Elia M, Connor CS, McGinness MK, Mammen JM, Wagner JL, Ward C, Ranallo L, Knight CJ, Stecklein SR, Jensen RA, Fabian CJ, Godwin AK. Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing. Breast Cancer Res Treat. 2014 Jun;145(3):707-14.

30. Greenblatt MS, Chappuis PO, Bond JP, Hamel N, Foulkes WD. TP53 mutations in breast cancer associated with BRCA1 or BRCA2 germ-line mutations: distinctive spectrum and structural distribution. Cancer Res. 2001;61:4092–7.

31. Zhang H, Somasundaram K, Peng Y, Tian H, Zhang H, Bi D, et al. BRCA1 physically associates with p53 and stimulates its transcriptional activity. Oncogene. 1998;16:1713–21.

32. Niyomnaitham S, Parinyanitikul N, Roothumnong E, Jinda W, Samarnthai N, Atikankul T, Suktitipat B, Thongnoppakhun W, Limwongse C, Pithukpakorn M. Tumor mutational profile of triple negative breast cancer patients in Thailand revealed distinctive genetic alteration in chromatin remodeling gene. PeerJ. 2019 Feb 25;7:e6501.

33. Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, Szwiec M, Byrski T, Debniak T, Gorski B, Sopik V, Akbari MR, Sun P, Gronwald J, Narod SA, Lubinski J, Polish Hereditary Breast Cancer Consortium Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. Lancet Oncol. 2015;16:638–644.

34. Neidhardt G, Hauke J, Ramser J, Groß E, Gehrig A, Müller CR, Kahlert AK, Hackmann K, Honisch E, Niederacher D, Heilmann-Heimbach S, Franke A, Lieb W, Thiele H, Altmüller J, Nürnberg P, Klaschik K, Ernst C, Ditsch N, Jessen F, Ramirez A, Wappenschmidt B, Engel C, Rhiem K, Meindl A, Schmutzler RK, Hahnen E. Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. JAMA Oncol. 2017 Sep 1;3(9):1245-1248.

35. Sun J, Meng H, Yao L, Lv M, Bai J, Zhang J, Wang L, Ouyang T, Li J, Wang T, et al. Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. Clin Cancer Res. 2017;23:6113–9.

36. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, *et al*. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol. 2015 Feb 1;33(4):304-11.

37. Shi, Y., Jin, J., Ji, W. *et al*. Therapeutic landscape in mutational triple negative breast cancer. Mol Cancer 17, 99 (2018).

38. Shargel L, Andrew B, Wu-Pong S (2005) Applied biopharmaceutics and pharmacokinetics. McGraw-Hill, New York.

39. Gonzalez-Angulo AM, Hennessy BT, Mills GB (2010) Future of personalized medicine in oncology: A systems biology approach. J Clin Onco. 28: 2777-2783.

40. Sacco K, Grech G. Actionable pharmacogenetic markers for prediction and prognosis in breast cancer. EPMA J. 2015 Jul 22;6(1):15.

41. Baldacchino S, Saliba C, Petroni V *et al*. (2014) Deregulation of the phosphatase, PP2A is a common event in breast cancer, predicting sensitivity to FTY720. EPMA J. 5(1):3.

42. Mandrekar SJ, Sargent DJ. (2009) Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol. 27(24):4027–34.

43. Jin Y, Desta Z, Stearns V *et al*. (2005) CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. J Natl Cancer Inst. 97(1):30–9.

44. Bradford LD (2002) CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics. 3(2):229–43.

45. Goetz MP, Knox SK, Suman VJ *et al*. (2007) The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. Breast Cancer Res Treat. 101(1):113-21.

46. Colomer R, Monzo M, Tusquets I *et al*. (2008) A single-nucleotide polymorphism in the aromatase gene is associated with the efficacy of the aromatase inhibitor letrozole in advanced breast carcinoma. Clin Cancer Res. 14(3):811–6.

47. Ma CX, Adjei AA, Salavaggione OE *et al*. (2005) Human aromatase: gene resequencing and functional genomics. Cancer Res. 65(23):11071–82.

48. Ingle JN. (2010) Genome-wide case-control study of musculoskeletal adverse events and functional genomics in women receiving aromatase inhibitors: going beyond associations. Breast Cancer Res. 10;12 Suppl 4: S17.

49. Hublin JJ, Ben-Ncer A, Bailey SE *et al*. (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature. 546:289–92.

50. Campbell MC, Hirbo JB, Townsend JP, *et al*. (2014) The peopling of the African continent and the diaspora into the new world. Curr Opin Genet Dev. 29:120–32.

51. Reich D, Thangaraj K, Patterson N, et al. (2009) Reconstructing Indian population history. Nature. 461(7263):489–94.

52. Majumder PP, Basu A (2014) A genomic view of the peopling and population structure of India. Cold Spring Harb Perspect Biol. 7(4):a008540.

53. Bamshad M, Kivisild T, Watkins WS, et al. (2001) Genetic evidence on the origins of Indian caste populations. Genome Res. 11(6):994–1004.

54. Kivisild T, Rootsi S, Metspalu M et al. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet. 72(2):313–32.

55. Das K, Malhotra KC, Mukherjee BN et al. (1996) Population structure and genetic differentiation among 16 tribal populations of Central India. Hum Biol. 68(5):679–705.

56. Majumder PP, Roy B, Banerjee S et al. (1999) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet. 7(4):435–46.

57. Thanseem I, Thangaraj K, Chaubey G et al. (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. BMC Genet. 7:42.

58. Indian Genome Variation Consortium (2005) The Indian Genome Variation database (IGVdb): a project overview. Hum. Genet. 118:1–11.

59. Jain A, Bhoyar RC, Pandhare K et al. (2020) IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. Nucleic Acids Res.

60. Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46:818–825.

61. 1000 Genomes Project Consortium, Auton A, Brooks LD et al. (2015) A global reference for human genetic variation. Nature. 526:68–74.

62. Gurdasani D, Carstensen T, Tekola-Ayele F et al. (2015) The African Genome Variation Project shapes medical genetics in Africa. Nature. 517:327–332.

63. Mallick S, Li H, Lipson M et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 538:201–206.

64. Nagasaki M, Yasuda J, Katsuoka F et al. (2015) Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat. Commun. 6:8018.

65. Scott EM, Halees A, Itan Y et al. (2016) Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat. Genet. 48:1071–1076.

66. Fakhro KA, Staudt MR, Ramstetter MD et al. (2016) The Qatar genome: a population-specific tool for precision medicine in the Middle East. Hum Genome Var. 3:16016.

67. Lek M, Karczewski KJ, Minikel EV et al. (2016) Analysis of protein-coding genetic variation in60,706 humans. Nature. 536:285–291.

68. Karczewski KJ, Francioli LC, Tiao G et al. (2019) Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 10.1101/531210.

69. Exome Variant Server (https://evs.gs.washington.edu/EVS/).

70. Sengupta D, Choudhury A, Basu A et al. (2016) Population Stratification and Underrepresentation of Indian Subcontinent Genetic Diversity in the 1000 Genomes Project Dataset. Genome Biol Evol. 8(11):3460-3470.

71. Indian Genome Variation Consortium (2008) Genetic landscape of the people of India: a canvas for disease gene exploration. J. Genet. 87:3–20.

72. Times of India: https://www.indiatimes.com/news/india/cancer-cases-on-the-rise-in-mizoram-state-being-called-cancer-capital-of-india-263190.html.

73. Prashant M, Krishnan S, Meesha C et al. (2020) Cancer Statistics, 2020: Report from National Cancer Registry Programme, India. JCO GO. 6:1063-1075.

74. Zomawia E, Zami Z, Vanlallawma A, Kumar NS, Zothanzama J, Tlau L, Chhakchhuak L, Pachuau L, Pautu JL, Hmangaihzuali EVL. Cancer awareness, diagnosis and treatment needs in Mizoram, India: evidence from 18 years trends (2003-2020). Lancet Reg Health Southeast Asia. 2023 Sep 21;17:100281.

75. Heng Li, Richard Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, Volume 25, Issue 14, July 2009, Pages 1754–1760.

76. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, Bioinformatics, Volume 25, Issue 16, August 2009, Pages 2078–2079,

77. http://broadinstitute.github.io/picard.

78. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data Nucleic Acids Research, 38:e164, 2010

79. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1.

80. Van der Auwera GA, O'Connor BD, (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. 1st ed. O'Reilly Media, California.

81. Karl Pearson F.R.S. . (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559–572.

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

83. Jiang Q, Greenberg RA. Deciphering the BRCA1 tumor suppressor network. J Biol Chem. 2015;290:17724–32.

84. Li ML, Greenberg RA. Links between genome integrity and BRCA1 tumor suppression. Trends Biochem Sci. 2012;37:418–24.

85. Kass EM, Moynahan ME, Jasin M. When genome maintenance goes badly Awry. Mol Cell. 2016;62:777–87.

86. King MC, Marks JH, Mandell JB, New York Breast Cancer Study, G. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003;302:643–6.

87. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161:1215–28.

88. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015;518:495–501.

89. Ichikawa H., Wakai T., Nagahashi M., Shimada Y., Hanyu T., Kanu Y., Muneoka Y., Ishikawa T., Takizawa K., Tajima Y., et al. Pathogenic germline BRCA1/2 mutations and familial predisposition to gastric cancer. JCO Precis. Oncol. 2018;2:18–00097.

90. Sahasrabudhe R., Lott P., Bohorquez M., Toal T., Estrada A.P., Suarez J.J., Brea-Fernandez A., Cameselle-Teijeiro J., Pinto C., Ramos I., et al. Germline Mutations in PALB2, BRCA1, and RAD51C, Which Regulate DNA Recombination Repair, in Patients with Gastric Cancer. Gastroenterology. 2017;152:983–986.e6.

91. Halpern N., Grinshpun A., Boursi B., Golan T., Margalit O., Aderka D., Friedman E., Laitman Y., Hubert A., Kadouri L., et al. Clinical Characteristics and Prognosis of Gastric Cancer Patients with BRCA 1/2 Germline Mutations: Report of Ten Cases and a Literature Review. Onco Targets Ther. 2020;13:11637–11644.

92. Correa P. Human gastric carcinogenesis: A multistep and multifactorial process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res. 1992;52:6735–6740.

93. Moiseyenko VM, Volkov NM, Suspistin EN, Yanus GA, Iyevleva AG, Kuligina ESh, Togo AV, Kornilov AV, Ivantsov AO, Imyanitov EN. Evidence for predictive role of BRCA1 and bTUBIII in gastric cancer. Med Oncol. 2013 Jun;30(2):545.

94. Avanesyan AA, Sokolenko AP, Ivantsov AO, e*t al*. Gastric Cancer in BRCA1 Germline Mutation Carriers: Results of Endoscopic Screening and Molecular Analysis of Tumor Tissues. Pathobiology. 2020;87(6):367-374.

95. Buckley KH, Niccum BA, Maxwell KN, Katona BW. Gastric Cancer Risk and Pathogenesis in BRCA1 and BRCA2 Carriers. Cancers (Basel). 2022 Dec 1;14(23):5953.

96. West AH, Knollman H, Dugan J, Hedeker D, Handorf EA, Nielsen SM, Bealin LC, Goldblatt LG, Willems H, Daly MB, Afghahi A, Olopade OI, Hulick PJ, Shagisultanova E, Huo D, Obeid E, Churpek JE. Hematologic toxicity in BRCA1 and BRCA2 mutation carriers during chemotherapy: A retrospective matched cohort study. Cancer Med. 2019 Sep;8(12):5609-5618.

97. Furlanetto J, Möbus V, Schneeweiss A, Rhiem K, Tesch H, Blohmer JU, Lübbe K, Untch M, Salat C, Huober J, Klare P, Schmutzler R, Couch FJ, Lederer B, Gerber B, Zahm DM, Bauerfeind I, Nekljudova V, Hanusch C, Jackisch C, Link T, Hahnen E, Loibl S, Fasching PA. Germline BRCA1/2 mutations and severe haematological toxicities in patients with breast cancer treated with neoadjuvant chemotherapy. Eur J Cancer. 2021 Mar;145:44-52.

98. Kumar S, Furrukh M, Al-Baimani K, Al-Ajmi A, Burney IA, Al-Moundhri MS. Outcomes of Women with Non-Metastatic Triple-Negative Breast Cancer in Oman: A single-centre experience. Sultan Qaboos Univ Med J. 2019 Aug;19(3).

99. Houvenaeghel G, Sabatier R, Reyal F, Classe JM, Giard S, Charitansky H, Rouzier R, Faure C, Garbay JR, Daraï E, Hudry D, Gimbergues P, Villet R, Lambaudie E. Axillary lymph node micrometastases decrease triple-negative early breast cancer survival. Br J Cancer. 2016 Oct 25;115(9):1024-1031.

100. Kaplun, A., Hogan, J., Schacherer, F. et al. PGMD: a comprehensive manually curated pharmacogenomic database. Pharmacogenomics J 16, 124–128 (2016).

101. Vatsis KP, Weber WW, Bell DA, Dupret JM, Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA, Relling MV, et al. Nomenclature for N-acetyltransferases. Pharmacogenetics. 1995 Feb;5(1):1-17.

102. Li D, Xie AH, Liu Z, Li D, Ning B, Thakkar S, Tong W, Xu J. Linking Pharmacogenomic Information on Drug Safety and Efficacy with Ethnic Minority Populations. Pharmaceutics. 2020 Oct 25;12(11):1021.

103. Bethke L., Webb E., Murray A., Schoemaker M., Feychting M., Lönn S., Ahlbom A., Malmer B., Henriksson R., Auvinen A., et al. Functional Polymorphisms in Folate Metabolism Genes Influence the Risk of Meningioma and Glioma. Cancer Epidemiol. Biomark. Prev. 2008;17:1195–1202.

104. Hasan T., Arora R., Bansal A.K., Bhattacharya R., Sharma G.S., Singh L.R. Disturbed homocysteine metabolism is associated with cancer. Exp. Mol. Med. 2019;51:1–13.
105. Bravatà V. Controversial roles of methylenetetrahydrofolate reductase polymorphisms and folate in breast cancer disease. Int. J. Food Sci. Nutr. 2014;66:43–49.
106. Matsuo K., Ito H., Wakai K., Hirose K., Saito T., Suzuki T., Kato T., Hirai T., Kanemitsu Y., Hamajima H., et al. One-carbon metabolism related gene polymorphisms interact with alcohol drinking to influence the risk of colorectal cancer in Japan. Carcinogenesis. 2005;26:2164–2171.
107. Singal R., Ferdinand L., Das P.M., Reis I.M., Schlesselman J.J. Polymorphisms in the methylenetetrahydrofolate reductase gene and prostate cancer risk. Int. J. Oncol. 2004;25:1465–1471.
108. Robien K., Ulrich C.M. 5,10-Methylenetetrahydrofolate reductase polymorphisms and leukemia risk: A HuGE minireview. Am. J. Epidemiol. 2003;157:571–582.
109. Tsang B.L., Devine O.J., Cordero A.M., Marchetta C.M., Mulinare J., Mersereau P., Guo J., Qi Y.P., Berry R.J., Rosenthal J., et al. Assessing the association between the methylenetetrahydrofolate reductase (MTHFR) 677C>T polymorphism and blood folate concentrations: A systematic review and meta-analysis of trials and observational studies. Am. J. Clin. Nutr. 2015;101:1286–1294.
110. Awwad N., Yousef A.-M., Abuhaliema A., Abdalla I., Yousef M. Relationship between Genetic Polymorphisms in MTHFR (C677T, A1298C and their Haplotypes) and the Incidence Of Breast Cancer among Jordanian Females—Case-Control Study. Asian Pac. J. Cancer Prev. 2015;16:5007–5011.
111. Hesari A., MalekSabet A., Tirkani A.N., Ghazizadeh H., Iranifar E., Mohagheg F., Anoshrvani A.A., Ghasemi F. Evaluation of the two polymorphisms rs1801133 in MTHFR and rs10811661 in CDKN2A/B in breast cancer. J. Cell. Biochem. 2019;120:2090–2097.
112. Meneses-Sanchez P., Garcia-Hernandez S.C., Porchia L.M., Pérez-Fuentes R., Torres-Rasgado E., Soto A.D.A., Gonzalez-Mejia M.E. C677T and A1298C

methylenetetrahydrofolate reductase polymorphisms and breast cancer susceptibility among Latinos: A meta-analysis. Breast Cancer. 2019;26:602–611.

113. He L., Shen Y. MTHFR C677T polymorphism and breast, ovarian cancer risk: A meta-analysis of 19,260 patients and 26,364 controls. Onco. Targets Ther. 2017;10:227–238.

114. Li K., Li W., Dong X. Association of 677 C>T (rs1801133) and 1298 A>C (rs1801131) Polymorphisms in the MTHFR Gene and Breast Cancer Susceptibility: A Meta-Analysis Based on 57 Individual Studies. PLoS ONE. 2014;9:e71290.

115. Liu W., Li Y., Li R., Han X., Ma Y., Liu B., Kong X. Association of mthfr A1298C polymorphism with breast cancer and/or ovarian cancer risk: An updated me-ta-analysis. Afr. J. Tradit. Complement. Altern. Med. 2016;13:72–86.

116. Ajaz S., Ali S.M., Siddiqa A., Memon M.A., Abid A., Khaliq S. Independent and Combined Associations of 677C/T and 1298A/C Polymorphisms in the MTHFR gene with Breast Cancers in a South-Asian Population. Medrxiv. 2021.

117. Durán M., Ángel C., Araújo M.D., Pinheiro D.D.R., Burbano R.M.R., Borges B.D.N. Thymidylate synthase and methylenetetrahydrofolate reductase polymorphisms and breast cancer susceptibility in a Brazilian population. Meta Gene. 2021;28:100889.

118. Houghton S.C., Eliassen A.H., Zhang S.M., Selhub J., Rosner B.A., Willett W.C., Hankinson S.E. Plasma B-vitamins and one-carbon metabolites and the risk of breast cancer in younger women. Breast Cancer Res. Treat. 2019;176:191–203.

119. Pooja S., Carlus J., Sekhar D., Francis A., Gupta N., Konwar R., Kumar S., Kumar S., Thangaraj K., Rajender S. MTHFR 677C>T Polymorphism and the Risk of Breast Cancer: Evidence from an Original Study and Pooled Data for 28031 Cases and 31880 Controls. PLoS ONE. 2015;10:e0120654.

120. Bravatà V. Controversial roles of methylenetetrahydrofolate reductase polymorphisms and folate in breast cancer disease. Int. J. Food Sci. Nutr. 2015;66:43–49.

121. Roth M, Obaidat A, Hagenbuch B. OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. BrJ Pharmacol. 2012;165(5):1260–1287.

122. Joyce H, McCann A, Clynes M, Larkin A. Influence of multidrug resistence and drug transport proteins on chemotherapy drug metabolism. Expert Opin Drug Metab Toxicol. 2015;11:795–809.

123. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol Rev. 2011;63:157–81.

124. Reimer T, Kempert S, Gerber B, Thiesen HJ, Hartmann S, Koczan D. SLCO1B1*5 polymorphism (rs4149056) is associated with chemotherapy-induced amenorrhea in premenopausal women with breast cancer: a prospective cohort study. BMC Cancer. 2016 May 27;16:337.

125. Shahrokni A, Rajebi MR, Saif MW. Toxicity and efficacy of 5-fluorouracil and capecitabine in a patient with TYMS gene polymorphism: a challenge or a dilemma?Clin. Colorectal Cancer 8(4), 231–234 (2009).

126. Scartozzi M, Maccaroni E, Giampieri R et al. 5-Fluorouracil pharmacogenomics: still rocking after all these years? Pharmacogenomics 12(2), 251–265 (2011).

127. Petak I, Tillman DM, Houghton JA, Lines C. p53 dependence of Fas induction and acute apoptosis in response to 5-fluorouracil-leucovorin in human colon carcinoma cell lines p53 dependence of Fas induction and acute apoptosis in response to 5-fluorouracil-leucovorin in human colon. Carcinoma 6, 4432–4441 (2000).

128. Anan K, Mitsuyama S, Tamae K et al. Increased dihydropyrimidine dehydrogenase activity in breast cancer. J. Surg. Oncol. 82(3),174–179 (2003).

129. van Kuilenburg AB, Dobritzsch D, Meinsma R, Haasjes J, Waterham HR, Nowaczyk MJ, Maropoulos GD, Hein G, Kalhoff H, Kirk JM, Baaske H, Aukett A, Duley JA, Ward KP, Lindqvist Y, van Gennip AH. Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. Biochem J. 2002 May 15;364(Pt 1):157-63.

130. Amstutz U, Henricks LM, Offer SM, Barbarino J, Schellens JHM, Swen JJ, Klein TE, McLeod HL, Caudle KE, Diasio RB, Schwab M. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for

dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. Clin Pharmacol Ther. 2018;103(2):210–216.

131. Vreken P, Van Kuilenburg AB, Meinsma R, Smit GP, Bakker HD, De Abreu RA, van Gennip AH. A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency. J Inherit Metab Dis. 1996;19(5):645–654.

132. Henricks LM, Lunenburg CA, Meulendijks D, Gelderblom H, Cats A, Swen JJ, Schellens JH, Guchelaar HJ. Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. Pharmacogenomics. 2015;16(11):1277–1286.

133. Hariprakash JM, Vellarikkal SK, Keechilat P, Verma A, Jayarajan R, Dixit V, Ravi R, Senthivel V, Kumar A, Sehgal P, Sonakar AK, Ambawat S, Giri AK, Philip A, Sivadas A, Faruq M, Bharadwaj D, Sivasubbu S, Scaria V. Pharmacogenetic landscape of DPYD variants in south Asian populations by integration of genome-scale data. Pharmacogenomics. 2018 Feb;19(3):227-241.

134. He YF, Wei W, Zhang X, Li YH, Li S, Wang FH, Lin XB, Li ZM, Zhang DS, Huang HQ, Hu B, Jiang WQ. Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in Chinese cancer patients. J Clin Pharm Ther. 2008 Jun;33(3):307-14.

135. Key TJ, Appleby PN, Reeves GK, et al. Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. J Natl Cancer Inst. 2003;95:1218–26.

136. Poutanen M, Isomaa V, Peltoketo H, Vihko R. Role of 17 beta-hydroxysteroid dehydrogenase type 1 in endocrine and intracrine estradiol biosynthesis. J Steroid Biochem Mol Biol. 1995;55:525–32.

137. Imaoka S, Yamada T, Hiroi T, et al. Multiple forms of human P450 expressed in Saccharomyces cerevisiae. Systematic characterization and comparison with those of the rat. Biochem Pharmacol. 1996;51:1041–50.

138. Apaya M. K., et al.., Integrated omics-based pathway analyses uncover CYP epoxygenase-associated networks as theranostic targets for metastatic triple negative breast cancer. J. Exp. Clin. Cancer Res. 38, 187 (2019).

139. Flanagan L., et al.., Efficacy of vitamin D compounds to modulate estrogen receptor negative breast cancer growth and invasion. J. Steroid Biochem. Mol. Biol. 84, 181–192 (2003).

140. Helvig C., Taimi M., Cameron D., Jones G., Petkovich M., Functional properties and substrate characterization of human CYP26A1, CYP26B1, and CYP26C1 expressed by recombinant baculovirus in insect cells. J. Pharmacol. Toxicol. Methods 64, 258–263 (2011).

141. Mangó, K., Fekete, F., Kiss, Á.F. et al. Association between CYP2B6 genetic variability and cyclophosphamide therapy in pediatric patients with neuroblastoma. Sci Rep 13, 11770 (2023).

142. Justenhoven C, Pentimalli D, Rabstein S, Harth V, Lotz A, Pesch B, Brüning T, Dörk T, Schürmann P, Bogdanova N, Park-Simon TW, Couch FJ, Olson JE, Fasching PA, Beckmann MW, Häberle L, Ekici A, Hall P, Czene K, Liu J, Li J, Baisch C, Hamann U, Ko YD, Brauch H. CYP2B6*6 is associated with increased breast cancer risk. Int J Cancer. 2014 Jan 15;134(2):426-30.

143. Bray J, Sludden J, Griffin MJ, Cole M, Verrill M, Jamieson D, Boddy AV. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. Br J Cancer. 2010 Mar 16;102(6):1003-9.

144. Kalra S, Kaur RP, Ludhiadch A, Shafi G, Vashista R, Kumar R, Munshi A. Association of CYP2C19*2 and ALDH1A1*1/*2 variants with disease outcome in breast cancer patients: results of a global screening array. Eur J Clin Pharmacol. 2018 Oct;74(10):1291-1298.

145. Cavanagh H, Rogers KMA. The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. Hered Cancer Clin Pract. 2015;13:16.

146. Mersch J, Jackson MA, Park M. et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer. 2015;121(2):269–275.

147. Jae Myoung Noh, Doo Ho Choi, Hyejin Baek. et al. Associations between BRCA Mutations in High-Risk Breast Cancer Patients and Familial Cancers Other than Breast or Ovary. J Breast Cancer. 2012;15(3):283–287.

148. Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V., and Boehning, D. (2012). Structure-Function of the Tumor Suppressor BRCA1. Comput. Struct. Biotechnol. J. 1, e201204005.

149. Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., et al. (2001). The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-Derived Mutation. J. Biol. Chem. 276 (18), 14537–14540.

150. Zhang, F., Fan, Q., Ren, K., and Andreassen, P. R. (2009). PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. Mol. Cancer Res. 7, 1110–1118.

151. Fu X, Tan W, Song Q, Pei H, Li J. BRCA1 and Breast Cancer: Molecular Mechanisms and Therapeutic Strategies. Front Cell Dev Biol. 2022 Mar 1;10:813457.

152. F. Wang, Q. Fang, Z. Ge, N. Yu, S. Xu, and X. Fan, "Common BRCA1 and BRCA2 mutations in breast cancer families: a meta-analysis from systematic review," Molecular Biology Reports, vol. 39, no. 3, pp. 2109–2118, 2012.

153. N. Tung, C. Battelli, B. Allen et al., "Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel," Cancer, vol. 121, no. 1, pp. 25–33, 2015.

154. Y. Zhu, J. Wu, C. Zhang et al., "BRCA mutations and survival in breast cancer: an updated systematic review and meta-analysis," Oncotarget, vol. 7, pp. 70113–70127, 2016.

155. J. Zhang, R. Pei, Z. Pang et al., "Prevalence and characterization of BRCA1 and BRCA2 germline mutations in Chinese women with familial breast cancer," Breast Cancer Research and Treatment, vol. 132, no. 2, pp. 421–428, 2012.

156. Fernandes VC, Golubeva VA, Di Pietro G, Shields C, Amankwah K, Nepomuceno TC, *et al.* Impact of amino acid substitutions at secondary structures in the BRCT domains of the tumor suppressor BRCA1: Implications for clinical annotation. J Biol Chem. 2019 Apr 12;294(15):5980-5992.

157. Petitalot A, Dardillac E, Jacquet E, Nhiri N, Guirouilh-Barbat J, Julien P, Bouazzaoui I, Bonte D, Feunteun J, Schnell JA, Lafitte P, Aude JC, Noguès C, Rouleau E, Lidereau R, Lopez BS, Zinn-Justin S, Caputo SM; UNICANCER

Genetic Group BRCA network. Combining Homologous Recombination and Phosphopeptide-binding Data to Predict the Impact of BRCA1 BRCT Variants on Cancer Risk. Mol Cancer Res. 2019 Jan;17(1):54-69.

158. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217-222.

159. Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, et al.. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. Hum Mutat. 2019 Sep;40(9):1557-1578.

160. Spina C., et al. Vitamin D and cancer. Anticancer Res. 2006;26(4A):2515–2524.

161. Jiang Y.J., et al. 1α, 25 (OH) 2-dihydroxyvitamin D3/VDR protects the skin from UVB-induced tumor formation by interacting with the β-catenin pathway. J. Steroid Biochem. Mol. Biol. 2013;136:229–232.

162. Feldman D, Krishnan AV, Swami S, Giovanucci E, Feldman BJ. The role of Vitamin D in reducing cancer risk and progression. Nat Rev. 2014;14:342–57.

163. Köstner K., et al. The relevance of vitamin D receptor (VDR) gene polymorphisms for cancer: a review of the literature. Anticancer Res. 2009;29(9):3511–3536.

164. Nam SE, Jin T, Park KS, Saindane M, Noh WC, Yoo YB, Park WS, Yun IJ. Vitamin D receptor (VDR) mRNA overexpression is associated with poor prognosis in breast carcinoma. Ann Surg Treat Res. 2022 Oct;103(4):183-194.

165. McCullough, M.L., Stevens, V.L., Diver, W.R. et al. Vitamin D pathway gene polymorphisms, diet, and risk of postmenopausal breast cancer: a nested case-control study. Breast Cancer Res 9, R9 (2007).

166. Akilzhanova A, Abilova Z, Sikhayeva N, Shtefanov I, Makishev A, Adylkhanov T, Rakhypbekov T, Zhumadilov Z, Momynaliev K. Vitamin D Receptor Gene Polymorphisms and Breast Cancer Risk in Kazakhstan. Cent Asian J Glob Health. 2014 Jan 24;2(Suppl):95.

167. Kazemi SM, Esmaieli-Bandboni A, Veisi Malekshahi Z, Shahbaz Sardood M, Hashemi M, Majidzadeh K, Kadkhodazadeh M, Esmaili R, Negahdari B. Vitamin D

receptor gene polymorphisms and risk of breast cancer in Iranian women. Ann Med Surg (Lond). 2021 Dec 4;73:103150.

168. Lurie G, Wilkens LR, Thompson PJ, Carney ME, Palmieri RT, Pharoah PD, Song H, Hogdall E, Kjaer SK, DiCioccio RA, McGuire V, Whittemore AS, Gayther SA, Gentry-Maharaj A, Menon U, Ramus SJ, Goodman MT; Ovarian Cancer Association Consortium. Vitamin D receptor rs2228570 polymorphism and invasive ovarian carcinoma risk: pooled analysis in five studies within the Ovarian Cancer Association Consortium. Int J Cancer. 2011 Feb 15;128(4):936-43.

169. Chen H, Zhu J. Vitamin D receptor rs2228570 polymorphism and susceptibility to ovarian cancer: An updated meta-analysis. J Obstet Gynaecol Res. 2018 Mar;44(3):556-565.

170. Uitterlinden AG, Fang Y, van Meurs JB, Pols HA, van Leeuwen JP. Genetics and biology of vitamin D receptor polymorphisms. Gene. 2004;338:143–156.

171. Terry SF (2015) Obama's Precision Medicine Initiative. Genet Test Mol Biomarkers. 19(3):113–114.

172. Mehrgou A, Akouchekian M (2016) The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. Med J Islam Repub Iran. 30:369.

173. Carol AP, Vincent C (2014) Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *J. Cancer Epidemiol.*

174. Onitilo AA, Engel JM, Greenlee RT et al. (2009) Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res. 7(1-2):4–13.

175. Collignon J, Lousberg L, Schroeder H et al. (2016) Triple-negative breast cancer: treatment challenges and solutions. Breast Cancer (Dove Med Press). 20(8):93-107.

176. Zaharia M, Gómez H. (2013) Cáncer de mama triple negativo: una enfermedad de difícil diagnóstico y tratamiento (Triple negative breast cancer: a difficult disease to diagnose and treat). Rev Peru Med Exp Salud Publica. 30(4):649-56.

177. Freda L, John Z, Vanlalhruaii, et al. (2020) Whole exome sequencing identifies the novel putative gene variants related with type 2 diabetes in Mizo population, northeast India. Gene.

178. Andrew V, Zothan Z, Jeremy LP et al. (2020) Pediatric leukemia could be driven predominantly by non-synonymous variants in mitochondrial complex V in Mizo population from Northeast India. Mitochondrial DNA A DNA Mapp Seq Anal. 31(6):245-249.

179. Vanlalhruaii TM, Vanlalbiakdiki SC, Zothansanga et al. (2020) Analysis of variants in mitochondrial genome and their putative pathogenicity in tuberculosis patients from Mizoram, North east India. Mitochondrion. 54:21-25.

180. C. S. Chakrabarti, M. Roy, N. K. Sengupta, R. Lalthantluanga, P. P. Majumder. Genetic relationships among some tribal groups inhabiting the north-eastern, eastern and sub-Himalayan regions of India. Ann. Hum. Genet. (2002), 66, 361–368.

181. Biswabandhu Bankura, Bishnupriya Basak, Prajjval Pratap Singh, Albert Vanlalruata, Arindam Chatterjee, Sudakshina Ghosh, Rakesh Tamang, Manjil Hazarika, Gyaneshwer Chaubey, Madhusudan Das. Northeast India: Genetic Inconsistency across Ethnicity and Geography. bioRxiv 2024.01.09.574778.

**BIBLOGRAPHY**


1. Hatton IA, Galbraith ED, Merleau NSC, Miettinen TP, Smith BM, Shander JA. The human cell count and size distribution. Proc Natl Acad Sci U S A. 2023 Sep 6;120(39).
2. Hanahan, D., & Weinberg, R. A. (2000). The hallmarks of cancer. Cell, 100(1), 57-70.
3. Bray, F., Ferlay, J., Soerjomataram, I., Siegel, R. L., Torre, L. A., & Jemal, A. (2018). Global cancer statistics 2018: GLOBOCAN estimates of incidence and mortality worldwide for 36 cancers in 185 countries. CA: A Cancer Journal for Clinicians, 68(6), 394-424.
4. Sung H, Ferlay J, Siegel RL, *et al*. Global Cancer Statistics 2020: GLOBOCAN Estimates of Incidence and Mortality Worldwide for 36 Cancers in 185 Countries. CA Cancer J Clin. 2021;71(3):209-249.
5. Henry NL, Shah PD, Haider I, Freer PE, Jagsi R, Sabel MS. Chapter 88: Cancer of the Breast. In: Niederhuber JE, Armitage JO, Doroshow JH, Kastan MB, Tepper JE, eds. Abeloff's Clinical Oncology. 6th ed. Philadelphia, Pa: Elsevier; 2020.
6. American Cancer Society. Accessed at https://www.cancer.org/cancer/types/breast-cancer/about/types-of-breast-cancer.htmlt on 10th October 2023.
7. Sharma GN, Dave R, Sanadya J, Sharma P, Sharma KK. Various types and management of breast cancer: an overview. J Adv Pharm Technol Res. 2010 Apr;1(2):109-26.
8. Perou CM, Sorlie T, Eisen MB, *et al*: Molecular portraits of human breast tumours. Nature 406:747-752, 2000.
9. Sotiriou C, Neo SY, McShane LM, *et al*. 2003. Breast cancer classification and prognosis based ongene expression profiles from a population-based study. Proc Natl Acad Sc i U S A 100:10393-10398.
10. Sørlie T, Perou CM, Tibshirani R, *et al*., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sc i U S A 98:10869-10874.

11. Sorlie T, Tibshirani R, Parker J, *et al*., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proc Natl Acad Sci U S A 100:8418-8423.

12. Zhao H, Langerod A, Ji Y, *et al*., 2004. Different gene expression patterns in invasive lobular and ductal carcinomas of the breast. Mol Biol Cell 15:2523-2536.

13. Calza S, Hall P, Auer G, *et al*. 2006. Intrinsic molecular signature of breast cancer in a population-based cohort of 412 patients. Breast Cancer Res 8: R34.

14. Hu Z, Fan C, Oh DS, *et al*. 2006. The molecular portraits of breast tumors are conserved across microarray platforms. BMC Genomics 7:96.

15. Rakha EA, Reis-Filho JS, Ellis IO. Basal-like breast cancer: a critical review. J Clin Oncol. 2008 May 20;26(15):2568-81.

16. Bergamaschi A, Kim YH, Wang P, et al: Distinct patterns of DNA copy number alteration are associated with different clinicopathological features and gene-expression subtypes of breast cancer. Genes Chromosomes Cancer 45:1033-1040, 2006

17. Chin K, DeVries S, Fridlyand J, et al: Genomic and transcriptional aberrations linked to breast cancer pathophysiologies. Cancer Cell 10:529-541, 2006

18. Vincent-Salomon A, Gruel N, Lucchesi C, et al: Identification of typical medullary breast carcinoma as a genomic sub-group of basal-like carcinomas, a heterogeneous new molecular entity. Breast Cancer Res 9:R24, 2007

19. Turner NC, Reis-Filho JS: Basal-like breast cancer and the BRCA1 phenotype. Oncogene 25:5846-5853, 2006

20. Reis-Filho JS, Pinheiro C, Lambros MB, et al: EGFR amplification and lack of activating mutations in metaplastic breast carcinomas. J Pathol 209:445-453, 2006.

21. Carey LA, Perou CM, Livasy CA, Dressler LG, Cowan D, Conway K, Karaca G, Troester MA, Tse CK, Edmiston S, Deming SL, Geradts J, Cheang MC, Nielsen TO, Moorman PG, Earp HS, Millikan RC. Race, breast cancer subtypes, and survival in the Carolina Breast Cancer Study. JAMA. 2006 Jun 7;295(21):2492-502.

22. Lehmann, B.D.; Bauer, J.A.; Chen, X.; Sanders, M.E.; Chakravarthy, A.B.; Shyr, Y.; Pietenpol, J.A. Identification of human triple-negative breast cancer

subtypes and preclinical models for selection of targeted therapies. J. Clin. Investig. 2011, 121,2750–2767.

23. Lehmann, B.D.; Jovanovic, B.; Chen, X.; Estrada, M.V.; Johnson, K.N.; Shyr, Y.; Moses, H.L.; Sanders, M.E.; Pietenpol, J.A. Refinement of Triple-Negative Breast Cancer Molecular Subtypes: Implications for Neoadjuvant Chemotherapy Selection. PLoS ONE 2016, 11.

24. Burstein, M.D.; Tsimelzon, A.; Poage, G.M.; Covington, K.R.; Contreras, A.; Fuqua, S.A.; Savage, M.I.; Osborne, C.K.; Hilsenbeck, S.G.; Chang, J.C.; et al. Comprehensive genomic analysis identifies novel subtypes and targets of triple-negative breast cancer. Clin. Cancer Res. 2015, 21, 1688–1698.

25. Jiang, Y.Z.; Liu, Y.; Xiao, Y.; Hu, X.; Jiang, L.; Zuo, W.J.; Ma, D.; Ding, J.; Zhu, X.; Zou, J.; et al. Molecular subtyping and genomic profiling expand precision medicine in refractory metastatic triple-negative breast cancer: The FUTURE trial. Cell Res. 2021, 31,178–186.

26. Stover, D.G.; Parsons, H.A.; Ha, G.; Freeman, S.S.; Barry, W.T.; Guo, H.; Choudhury, A.D.; Gydush, G.; Reed, S.C.; Rhoades, J.; et al. Association of Cell-Free DNA Tumor Fraction and Somatic Copy Number Alterations With Survival in Metastatic Triple-Negative Breast Cancer. J. Clin. Oncol. 2018, 36, 543–553. [CrossRef] [PubMed]

27. Shah, S.P.; Roth, A.; Goya, R.; Oloumi, A.; Ha, G.; Zhao, Y.; Turashvili, G.; Ding, J.; Tse, K.; Haffari, G.; et al. The clonal and mutational evolution spectrum of primary triple-negative breast cancers. Nature 2012, 486, 395–399.

28. Cancer Genome Atlas, N. Comprehensive molecular portraits of human breast tumours. Nature 2012, 490, 61–70.

29. Sharma P, Klemp JR, Kimler BF, Mahnken JD, Geier LJ, Khan QJ, Elia M, Connor CS, McGinness MK, Mammen JM, Wagner JL, Ward C, Ranallo L, Knight CJ, Stecklein SR, Jensen RA, Fabian CJ, Godwin AK. Germline BRCA mutation evaluation in a prospective triple-negative breast cancer registry: implications for hereditary breast and/or ovarian cancer syndrome testing. Breast Cancer Res Treat. 2014 Jun;145(3):707-14.

30. Greenblatt MS, Chappuis PO, Bond JP, Hamel N, Foulkes WD. TP53 mutations in breast cancer associated with BRCA1 or BRCA2 germ-line mutations: distinctive spectrum and structural distribution. Cancer Res. 2001;61:4092–7.

31. Zhang H, Somasundaram K, Peng Y, Tian H, Zhang H, Bi D, et al. BRCA1 physically associates with p53 and stimulates its transcriptional activity. Oncogene. 1998;16:1713–21.

32. Niyomnaitham S, Parinyanitikul N, Roothumnong E, Jinda W, Samarnthai N, Atikankul T, Suktitipat B, Thongnoppakhun W, Limwongse C, Pithukpakorn M. Tumor mutational profile of triple negative breast cancer patients in Thailand revealed distinctive genetic alteration in chromatin remodeling gene. PeerJ. 2019 Feb 25;7:e6501.

33. Cybulski C, Kluzniak W, Huzarski T, Wokolorczyk D, Kashyap A, Jakubowska A, Szwiec M, Byrski T, Debniak T, Gorski B, Sopik V, Akbari MR, Sun P, Gronwald J, Narod SA, Lubinski J, Polish Hereditary Breast Cancer Consortium Clinical outcomes in women with breast cancer and a PALB2 mutation: a prospective cohort analysis. Lancet Oncol. 2015;16:638–644.

34. Neidhardt G, Hauke J, Ramser J, Groß E, Gehrig A, Müller CR, Kahlert AK, Hackmann K, Honisch E, Niederacher D, Heilmann-Heimbach S, Franke A, Lieb W, Thiele H, Altmüller J, Nürnberg P, Klaschik K, Ernst C, Ditsch N, Jessen F, Ramirez A, Wappenschmidt B, Engel C, Rhiem K, Meindl A, Schmutzler RK, Hahnen E. Association Between Loss-of-Function Mutations Within the FANCM Gene and Early-Onset Familial Breast Cancer. JAMA Oncol. 2017 Sep 1;3(9):1245-1248.

35. Sun J, Meng H, Yao L, Lv M, Bai J, Zhang J, Wang L, Ouyang T, Li J, Wang T, et al. Germline mutations in cancer susceptibility genes in a large series of unselected breast cancer patients. Clin Cancer Res. 2017;23:6113–9.

36. Couch FJ, Hart SN, Sharma P, Toland AE, Wang X, *et al*. Inherited mutations in 17 breast cancer susceptibility genes among a large triple-negative breast cancer cohort unselected for family history of breast cancer. J Clin Oncol. 2015 Feb 1;33(4):304-11.

37. Shi, Y., Jin, J., Ji, W. *et al*. Therapeutic landscape in mutational triple negative breast cancer. Mol Cancer 17, 99 (2018).

38. Shargel L, Andrew B, Wu-Pong S (2005) Applied biopharmaceutics and pharmacokinetics. McGraw-Hill, New York.

39. Gonzalez-Angulo AM, Hennessy BT, Mills GB (2010) Future of personalized medicine in oncology: A systems biology approach. J Clin Onco. 28: 2777-2783.

40. Sacco K, Grech G. Actionable pharmacogenetic markers for prediction and prognosis in breast cancer. EPMA J. 2015 Jul 22;6(1):15.

41. Baldacchino S, Saliba C, Petroni V *et al*. (2014) Deregulation of the phosphatase, PP2A is a common event in breast cancer, predicting sensitivity to FTY720. EPMA J. 5(1):3.

42. Mandrekar SJ, Sargent DJ. (2009) Clinical trial designs for predictive biomarker validation: theoretical considerations and practical challenges. J Clin Oncol. 27(24):4027–34.

43. Jin Y, Desta Z, Stearns V *et al*. (2005) CYP2D6 genotype, antidepressant use, and tamoxifen metabolism during adjuvant breast cancer treatment. J Natl Cancer Inst. 97(1):30–9.

44. Bradford LD (2002) CYP2D6 allele frequency in European Caucasians, Asians, Africans and their descendants. Pharmacogenomics. 3(2):229–43.

45. Goetz MP, Knox SK, Suman VJ *et al*. (2007) The impact of cytochrome P450 2D6 metabolism in women receiving adjuvant tamoxifen. Breast Cancer Res Treat. 101(1):113-21.

46. Colomer R, Monzo M, Tusquets I *et al*. (2008) A single-nucleotide polymorphism in the aromatase gene is associated with the efficacy of the aromatase inhibitor letrozole in advanced breast carcinoma. Clin Cancer Res. 14(3):811–6.

47. Ma CX, Adjei AA, Salavaggione OE *et al*. (2005) Human aromatase: gene resequencing and functional genomics. Cancer Res. 65(23):11071–82.

48. Ingle JN. (2010) Genome-wide case-control study of musculoskeletal adverse events and functional genomics in women receiving aromatase inhibitors: going beyond associations. Breast Cancer Res. 10;12 Suppl 4: S17.

49. Hublin JJ, Ben-Ncer A, Bailey SE *et al*. (2017) New fossils from Jebel Irhoud, Morocco and the pan-African origin of Homo sapiens. Nature. 546:289–92.

50. Campbell MC, Hirbo JB, Townsend JP, *et al*. (2014) The peopling of the African continent and the diaspora into the new world. Curr Opin Genet Dev. 29:120–32.

51. Reich D, Thangaraj K, Patterson N, et al. (2009) Reconstructing Indian population history. Nature. 461(7263):489–94.

52. Majumder PP, Basu A (2014) A genomic view of the peopling and population structure of India. Cold Spring Harb Perspect Biol. 7(4):a008540.

53. Bamshad M, Kivisild T, Watkins WS, et al. (2001) Genetic evidence on the origins of Indian caste populations. Genome Res. 11(6):994–1004.

54. Kivisild T, Rootsi S, Metspalu M et al. (2003) The genetic heritage of the earliest settlers persists both in Indian tribal and caste populations. Am J Hum Genet. 72(2):313–32.

55. Das K, Malhotra KC, Mukherjee BN et al. (1996) Population structure and genetic differentiation among 16 tribal populations of Central India. Hum Biol. 68(5):679–705.

56. Majumder PP, Roy B, Banerjee S et al. (1999) Human-specific insertion/deletion polymorphisms in Indian populations and their possible evolutionary implications. Eur J Hum Genet. 7(4):435–46.

57. Thanseem I, Thangaraj K, Chaubey G et al. (2006) Genetic affinities among the lower castes and tribal groups of India: inference from Y chromosome and mitochondrial DNA. BMC Genet. 7:42.

58. Indian Genome Variation Consortium (2005) The Indian Genome Variation database (IGVdb): a project overview. Hum. Genet. 118:1–11.

59. Jain A, Bhoyar RC, Pandhare K et al. (2020) IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. Nucleic Acids Res.

60. Genome of the Netherlands Consortium (2014) Whole-genome sequence variation, population structure and demographic history of the Dutch population. Nat. Genet. 46:818–825.

61. 1000 Genomes Project Consortium, Auton A, Brooks LD et al. (2015) A global reference for human genetic variation. Nature. 526:68–74.

62. Gurdasani D, Carstensen T, Tekola-Ayele F et al. (2015) The African Genome Variation Project shapes medical genetics in Africa. Nature. 517:327–332.

63. Mallick S, Li H, Lipson M et al. (2016) The Simons Genome Diversity Project: 300 genomes from 142 diverse populations. Nature. 538:201–206.

64. Nagasaki M, Yasuda J, Katsuoka F et al. (2015) Rare variant discovery by deep whole-genome sequencing of 1,070 Japanese individuals. Nat. Commun. 6:8018.

65. Scott EM, Halees A, Itan Y et al. (2016) Characterization of Greater Middle Eastern genetic variation for enhanced disease gene discovery. Nat. Genet. 48:1071–1076.

66. Fakhro KA, Staudt MR, Ramstetter MD et al. (2016) The Qatar genome: a population-specific tool for precision medicine in the Middle East. Hum Genome Var. 3:16016.

67. Lek M, Karczewski KJ, Minikel EV et al. (2016) Analysis of protein-coding genetic variation in60,706 humans. Nature. 536:285–291.

68. Karczewski KJ, Francioli LC, Tiao G et al. (2019) Variation across 141,456 human exomes and genomes reveals the spectrum of loss-of-function intolerance across human protein-coding genes. bioRxiv. 10.1101/531210.

69. Exome Variant Server (https://evs.gs.washington.edu/EVS/).

70. Sengupta D, Choudhury A, Basu A et al. (2016) Population Stratification and Underrepresentation of Indian Subcontinent Genetic Diversity in the 1000 Genomes Project Dataset. Genome Biol Evol. 8(11):3460-3470.

71. Indian Genome Variation Consortium (2008) Genetic landscape of the people of India: a canvas for disease gene exploration. J. Genet. 87:3–20.

72. Times of India: https://www.indiatimes.com/news/india/cancer-cases-on-the-rise-in-mizoram-state-being-called-cancer-capital-of-india-263190.html.

73. Prashant M, Krishnan S, Meesha C et al. (2020) Cancer Statistics, 2020: Report from National Cancer Registry Programme, India. JCO GO. 6:1063-1075.

74. Zomawia E, Zami Z, Vanlallawma A, Kumar NS, Zothanzama J, Tlau L, Chhakchhuak L, Pachuau L, Pautu JL, Hmangaihzuali EVL. Cancer awareness, diagnosis and treatment needs in Mizoram, India: evidence from 18 years trends (2003-2020). Lancet Reg Health Southeast Asia. 2023 Sep 21;17:100281.

75. Heng Li, Richard Durbin, Fast and accurate short read alignment with Burrows–Wheeler transform, Bioinformatics, Volume 25, Issue 14, July 2009, Pages 1754–1760.

76. Heng Li, Bob Handsaker, Alec Wysoker, Tim Fennell, Jue Ruan, Nils Homer, Gabor Marth, Goncalo Abecasis, Richard Durbin, 1000 Genome Project Data Processing Subgroup, The Sequence Alignment/Map format and SAMtools, Bioinformatics, Volume 25, Issue 16, August 2009, Pages 2078–2079,

77. http://broadinstitute.github.io/picard.

78. Wang K, Li M, Hakonarson H. ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data Nucleic Acids Research, 38:e164, 2010

79. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014 Aug 1;30(15):2114-20. doi: 10.1093/bioinformatics/btu170. Epub 2014 Apr 1.

80. Van der Auwera GA, O'Connor BD, (2020). Genomics in the Cloud: Using Docker, GATK, and WDL in Terra. 1st ed. O'Reilly Media, California.

81. Karl Pearson F.R.S. . (1901). LIII. On lines and planes of closest fit to systems of points in space. The London, Edinburgh, and Dublin Philosophical Magazine and Journal of Science, 2(11), 559–572.

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

83. Jiang Q, Greenberg RA. Deciphering the BRCA1 tumor suppressor network. J Biol Chem. 2015;290:17724–32.

84. Li ML, Greenberg RA. Links between genome integrity and BRCA1 tumor suppression. Trends Biochem Sci. 2012;37:418–24.

85. Kass EM, Moynahan ME, Jasin M. When genome maintenance goes badly Awry. Mol Cell. 2016;62:777–87.

86. King MC, Marks JH, Mandell JB, New York Breast Cancer Study, G. Breast and ovarian cancer risks due to inherited mutations in BRCA1 and BRCA2. Science. 2003;302:643–6.

87. Robinson D, Van Allen EM, Wu YM, Schultz N, Lonigro RJ, Mosquera JM, Montgomery B, Taplin ME, Pritchard CC, Attard G, et al. Integrative clinical genomics of advanced prostate cancer. Cell. 2015;161:1215–28.

88. Waddell N, Pajic M, Patch AM, Chang DK, Kassahn KS, Bailey P, Johns AL, Miller D, Nones K, Quek K, et al. Whole genomes redefine the mutational landscape of pancreatic cancer. Nature. 2015;518:495–501.

89. Ichikawa H., Wakai T., Nagahashi M., Shimada Y., Hanyu T., Kanu Y., Muneoka Y., Ishikawa T., Takizawa K., Tajima Y., et al. Pathogenic germline BRCA1/2 mutations and familial predisposition to gastric cancer. JCO Precis. Oncol. 2018;2:18–00097.

90. Sahasrabudhe R., Lott P., Bohorquez M., Toal T., Estrada A.P., Suarez J.J., Brea-Fernandez A., Cameselle-Teijeiro J., Pinto C., Ramos I., et al. Germline Mutations in PALB2, BRCA1, and RAD51C, Which Regulate DNA Recombination Repair, in Patients with Gastric Cancer. Gastroenterology. 2017;152:983–986.e6.

91. Halpern N., Grinshpun A., Boursi B., Golan T., Margalit O., Aderka D., Friedman E., Laitman Y., Hubert A., Kadouri L., et al. Clinical Characteristics and Prognosis of Gastric Cancer Patients with BRCA 1/2 Germline Mutations: Report of Ten Cases and a Literature Review. Onco Targets Ther. 2020;13:11637–11644.

92. Correa P. Human gastric carcinogenesis: A multistep and multifactorial process—First American Cancer Society Award Lecture on Cancer Epidemiology and Prevention. Cancer Res. 1992;52:6735–6740.

93. Moiseyenko VM, Volkov NM, Suspistin EN, Yanus GA, Iyevleva AG, Kuligina ESh, Togo AV, Kornilov AV, Ivantsov AO, Imyanitov EN. Evidence for predictive role of BRCA1 and bTUBIII in gastric cancer. Med Oncol. 2013 Jun;30(2):545.

94. Avanesyan AA, Sokolenko AP, Ivantsov AO, e*t al*. Gastric Cancer in BRCA1 Germline Mutation Carriers: Results of Endoscopic Screening and Molecular Analysis of Tumor Tissues. Pathobiology. 2020;87(6):367-374.

95. Buckley KH, Niccum BA, Maxwell KN, Katona BW. Gastric Cancer Risk and Pathogenesis in BRCA1 and BRCA2 Carriers. Cancers (Basel). 2022 Dec 1;14(23):5953.

96. West AH, Knollman H, Dugan J, Hedeker D, Handorf EA, Nielsen SM, Bealin LC, Goldblatt LG, Willems H, Daly MB, Afghahi A, Olopade OI, Hulick PJ, Shagisultanova E, Huo D, Obeid E, Churpek JE. Hematologic toxicity in BRCA1 and BRCA2 mutation carriers during chemotherapy: A retrospective matched cohort study. Cancer Med. 2019 Sep;8(12):5609-5618.

97. Furlanetto J, Möbus V, Schneeweiss A, Rhiem K, Tesch H, Blohmer JU, Lübbe K, Untch M, Salat C, Huober J, Klare P, Schmutzler R, Couch FJ, Lederer B, Gerber B, Zahm DM, Bauerfeind I, Nekljudova V, Hanusch C, Jackisch C, Link T, Hahnen E, Loibl S, Fasching PA. Germline BRCA1/2 mutations and severe haematological toxicities in patients with breast cancer treated with neoadjuvant chemotherapy. Eur J Cancer. 2021 Mar;145:44-52.

98. Kumar S, Furrukh M, Al-Baimani K, Al-Ajmi A, Burney IA, Al-Moundhri MS. Outcomes of Women with Non-Metastatic Triple-Negative Breast Cancer in Oman: A single-centre experience. Sultan Qaboos Univ Med J. 2019 Aug;19(3).

99. Houvenaeghel G, Sabatier R, Reyal F, Classe JM, Giard S, Charitansky H, Rouzier R, Faure C, Garbay JR, Daraï E, Hudry D, Gimbergues P, Villet R, Lambaudie E. Axillary lymph node micrometastases decrease triple-negative early breast cancer survival. Br J Cancer. 2016 Oct 25;115(9):1024-1031.

100. Kaplun, A., Hogan, J., Schacherer, F. et al. PGMD: a comprehensive manually curated pharmacogenomic database. Pharmacogenomics J 16, 124–128 (2016).

101. Vatsis KP, Weber WW, Bell DA, Dupret JM, Evans DA, Grant DM, Hein DW, Lin HJ, Meyer UA, Relling MV, et al. Nomenclature for N-acetyltransferases. Pharmacogenetics. 1995 Feb;5(1):1-17.

102. Li D, Xie AH, Liu Z, Li D, Ning B, Thakkar S, Tong W, Xu J. Linking Pharmacogenomic Information on Drug Safety and Efficacy with Ethnic Minority Populations. Pharmaceutics. 2020 Oct 25;12(11):1021.

103. Bethke L., Webb E., Murray A., Schoemaker M., Feychting M., Lönn S., Ahlbom A., Malmer B., Henriksson R., Auvinen A., et al. Functional Polymorphisms in Folate Metabolism Genes Influence the Risk of Meningioma and Glioma. Cancer Epidemiol. Biomark. Prev. 2008;17:1195–1202.

104. Hasan T., Arora R., Bansal A.K., Bhattacharya R., Sharma G.S., Singh L.R. Disturbed homocysteine metabolism is associated with cancer. Exp. Mol. Med. 2019;51:1–13.
105. Bravatà V. Controversial roles of methylenetetrahydrofolate reductase polymorphisms and folate in breast cancer disease. Int. J. Food Sci. Nutr. 2014;66:43–49.
106. Matsuo K., Ito H., Wakai K., Hirose K., Saito T., Suzuki T., Kato T., Hirai T., Kanemitsu Y., Hamajima H., et al. One-carbon metabolism related gene polymorphisms interact with alcohol drinking to influence the risk of colorectal cancer in Japan. Carcinogenesis. 2005;26:2164–2171.
107. Singal R., Ferdinand L., Das P.M., Reis I.M., Schlesselman J.J. Polymorphisms in the methylenetetrahydrofolate reductase gene and prostate cancer risk. Int. J. Oncol. 2004;25:1465–1471.
108. Robien K., Ulrich C.M. 5,10-Methylenetetrahydrofolate reductase polymorphisms and leukemia risk: A HuGE minireview. Am. J. Epidemiol. 2003;157:571–582.
109. Tsang B.L., Devine O.J., Cordero A.M., Marchetta C.M., Mulinare J., Mersereau P., Guo J., Qi Y.P., Berry R.J., Rosenthal J., et al. Assessing the association between the methylenetetrahydrofolate reductase (MTHFR) 677C>T polymorphism and blood folate concentrations: A systematic review and meta-analysis of trials and observational studies. Am. J. Clin. Nutr. 2015;101:1286–1294.
110. Awwad N., Yousef A.-M., Abuhaliema A., Abdalla I., Yousef M. Relationship between Genetic Polymorphisms in MTHFR (C677T, A1298C and their Haplotypes) and the Incidence Of Breast Cancer among Jordanian Females—Case-Control Study. Asian Pac. J. Cancer Prev. 2015;16:5007–5011.
111. Hesari A., MalekSabet A., Tirkani A.N., Ghazizadeh H., Iranifar E., Mohagheg F., Anoshrvani A.A., Ghasemi F. Evaluation of the two polymorphisms rs1801133 in MTHFR and rs10811661 in CDKN2A/B in breast cancer. J. Cell. Biochem. 2019;120:2090–2097.
112. Meneses-Sanchez P., Garcia-Hernandez S.C., Porchia L.M., Pérez-Fuentes R., Torres-Rasgado E., Soto A.D.A., Gonzalez-Mejia M.E. C677T and A1298C

methylenetetrahydrofolate reductase polymorphisms and breast cancer susceptibility among Latinos: A meta-analysis. Breast Cancer. 2019;26:602–611.

113. He L., Shen Y. MTHFR C677T polymorphism and breast, ovarian cancer risk: A meta-analysis of 19,260 patients and 26,364 controls. Onco. Targets Ther. 2017;10:227–238.

114. Li K., Li W., Dong X. Association of 677 C>T (rs1801133) and 1298 A>C (rs1801131) Polymorphisms in the MTHFR Gene and Breast Cancer Susceptibility: A Meta-Analysis Based on 57 Individual Studies. PLoS ONE. 2014;9:e71290.

115. Liu W., Li Y., Li R., Han X., Ma Y., Liu B., Kong X. Association of mthfr A1298C polymorphism with breast cancer and/or ovarian cancer risk: An updated me-ta-analysis. Afr. J. Tradit. Complement. Altern. Med. 2016;13:72–86.

116. Ajaz S., Ali S.M., Siddiqa A., Memon M.A., Abid A., Khaliq S. Independent and Combined Associations of 677C/T and 1298A/C Polymorphisms in the MTHFR gene with Breast Cancers in a South-Asian Population. Medrxiv. 2021.

117. Durán M., Ángel C., Araújo M.D., Pinheiro D.D.R., Burbano R.M.R., Borges B.D.N. Thymidylate synthase and methylenetetrahydrofolate reductase polymorphisms and breast cancer susceptibility in a Brazilian population. Meta Gene. 2021;28:100889.

118. Houghton S.C., Eliassen A.H., Zhang S.M., Selhub J., Rosner B.A., Willett W.C., Hankinson S.E. Plasma B-vitamins and one-carbon metabolites and the risk of breast cancer in younger women. Breast Cancer Res. Treat. 2019;176:191–203.

119. Pooja S., Carlus J., Sekhar D., Francis A., Gupta N., Konwar R., Kumar S., Kumar S., Thangaraj K., Rajender S. MTHFR 677C>T Polymorphism and the Risk of Breast Cancer: Evidence from an Original Study and Pooled Data for 28031 Cases and 31880 Controls. PLoS ONE. 2015;10:e0120654.

120. Bravatà V. Controversial roles of methylenetetrahydrofolate reductase polymorphisms and folate in breast cancer disease. Int. J. Food Sci. Nutr. 2015;66:43–49.

121. Roth M, Obaidat A, Hagenbuch B. OATPs, OATs and OCTs: the organic anion and cation transporters of the SLCO and SLC22A gene superfamilies. BrJ Pharmacol. 2012;165(5):1260–1287.

122. Joyce H, McCann A, Clynes M, Larkin A. Influence of multidrug resistence and drug transport proteins on chemotherapy drug metabolism. Expert Opin Drug Metab Toxicol. 2015;11:795–809.

123. Niemi M, Pasanen MK, Neuvonen PJ. Organic anion transporting polypeptide 1B1: a genetically polymorphic transporter of major importance for hepatic drug uptake. Pharmacol Rev. 2011;63:157–81.

124. Reimer T, Kempert S, Gerber B, Thiesen HJ, Hartmann S, Koczan D. SLCO1B1*5 polymorphism (rs4149056) is associated with chemotherapy-induced amenorrhea in premenopausal women with breast cancer: a prospective cohort study. BMC Cancer. 2016 May 27;16:337.

125. Shahrokni A, Rajebi MR, Saif MW. Toxicity and efficacy of 5-fluorouracil and capecitabine in a patient with TYMS gene polymorphism: a challenge or a dilemma?Clin. Colorectal Cancer 8(4), 231–234 (2009).

126. Scartozzi M, Maccaroni E, Giampieri R et al. 5-Fluorouracil pharmacogenomics: still rocking after all these years? Pharmacogenomics 12(2), 251–265 (2011).

127. Petak I, Tillman DM, Houghton JA, Lines C. p53 dependence of Fas induction and acute apoptosis in response to 5-fluorouracil-leucovorin in human colon carcinoma cell lines p53 dependence of Fas induction and acute apoptosis in response to 5-fluorouracil-leucovorin in human colon. Carcinoma 6, 4432–4441 (2000).

128. Anan K, Mitsuyama S, Tamae K et al. Increased dihydropyrimidine dehydrogenase activity in breast cancer. J. Surg. Oncol. 82(3),174–179 (2003).

129. van Kuilenburg AB, Dobritzsch D, Meinsma R, Haasjes J, Waterham HR, Nowaczyk MJ, Maropoulos GD, Hein G, Kalhoff H, Kirk JM, Baaske H, Aukett A, Duley JA, Ward KP, Lindqvist Y, van Gennip AH. Novel disease-causing mutations in the dihydropyrimidine dehydrogenase gene interpreted by analysis of the three-dimensional protein structure. Biochem J. 2002 May 15;364(Pt 1):157-63.

130. Amstutz U, Henricks LM, Offer SM, Barbarino J, Schellens JHM, Swen JJ, Klein TE, McLeod HL, Caudle KE, Diasio RB, Schwab M. Clinical Pharmacogenetics Implementation Consortium (CPIC) guideline for

dihydropyrimidine dehydrogenase genotype and fluoropyrimidine dosing: 2017 update. Clin Pharmacol Ther. 2018;103(2):210–216.

131. Vreken P, Van Kuilenburg AB, Meinsma R, Smit GP, Bakker HD, De Abreu RA, van Gennip AH. A point mutation in an invariant splice donor site leads to exon skipping in two unrelated Dutch patients with dihydropyrimidine dehydrogenase deficiency. J Inherit Metab Dis. 1996;19(5):645–654.

132. Henricks LM, Lunenburg CA, Meulendijks D, Gelderblom H, Cats A, Swen JJ, Schellens JH, Guchelaar HJ. Translating DPYD genotype into DPD phenotype: using the DPYD gene activity score. Pharmacogenomics. 2015;16(11):1277–1286.

133. Hariprakash JM, Vellarikkal SK, Keechilat P, Verma A, Jayarajan R, Dixit V, Ravi R, Senthivel V, Kumar A, Sehgal P, Sonakar AK, Ambawat S, Giri AK, Philip A, Sivadas A, Faruq M, Bharadwaj D, Sivasubbu S, Scaria V. Pharmacogenetic landscape of DPYD variants in south Asian populations by integration of genome-scale data. Pharmacogenomics. 2018 Feb;19(3):227-241.

134. He YF, Wei W, Zhang X, Li YH, Li S, Wang FH, Lin XB, Li ZM, Zhang DS, Huang HQ, Hu B, Jiang WQ. Analysis of the DPYD gene implicated in 5-fluorouracil catabolism in Chinese cancer patients. J Clin Pharm Ther. 2008 Jun;33(3):307-14.

135. Key TJ, Appleby PN, Reeves GK, et al. Body mass index, serum sex hormones, and breast cancer risk in postmenopausal women. J Natl Cancer Inst. 2003;95:1218–26.

136. Poutanen M, Isomaa V, Peltoketo H, Vihko R. Role of 17 beta-hydroxysteroid dehydrogenase type 1 in endocrine and intracrine estradiol biosynthesis. J Steroid Biochem Mol Biol. 1995;55:525–32.

137. Imaoka S, Yamada T, Hiroi T, et al. Multiple forms of human P450 expressed in Saccharomyces cerevisiae. Systematic characterization and comparison with those of the rat. Biochem Pharmacol. 1996;51:1041–50.

138. Apaya M. K., et al.., Integrated omics-based pathway analyses uncover CYP epoxygenase-associated networks as theranostic targets for metastatic triple negative breast cancer. J. Exp. Clin. Cancer Res. 38, 187 (2019).

139. Flanagan L., et al.., Efficacy of vitamin D compounds to modulate estrogen receptor negative breast cancer growth and invasion. J. Steroid Biochem. Mol. Biol. 84, 181–192 (2003).

140. Helvig C., Taimi M., Cameron D., Jones G., Petkovich M., Functional properties and substrate characterization of human CYP26A1, CYP26B1, and CYP26C1 expressed by recombinant baculovirus in insect cells. J. Pharmacol. Toxicol. Methods 64, 258–263 (2011).

141. Mangó, K., Fekete, F., Kiss, Á.F. et al. Association between CYP2B6 genetic variability and cyclophosphamide therapy in pediatric patients with neuroblastoma. Sci Rep 13, 11770 (2023).

142. Justenhoven C, Pentimalli D, Rabstein S, Harth V, Lotz A, Pesch B, Brüning T, Dörk T, Schürmann P, Bogdanova N, Park-Simon TW, Couch FJ, Olson JE, Fasching PA, Beckmann MW, Häberle L, Ekici A, Hall P, Czene K, Liu J, Li J, Baisch C, Hamann U, Ko YD, Brauch H. CYP2B6*6 is associated with increased breast cancer risk. Int J Cancer. 2014 Jan 15;134(2):426-30.

143. Bray J, Sludden J, Griffin MJ, Cole M, Verrill M, Jamieson D, Boddy AV. Influence of pharmacogenetics on response and toxicity in breast cancer patients treated with doxorubicin and cyclophosphamide. Br J Cancer. 2010 Mar 16;102(6):1003-9.

144. Kalra S, Kaur RP, Ludhiadch A, Shafi G, Vashista R, Kumar R, Munshi A. Association of CYP2C19*2 and ALDH1A1*1/*2 variants with disease outcome in breast cancer patients: results of a global screening array. Eur J Clin Pharmacol. 2018 Oct;74(10):1291-1298.

145. Cavanagh H, Rogers KMA. The role of BRCA1 and BRCA2 mutations in prostate, pancreatic and stomach cancers. Hered Cancer Clin Pract. 2015;13:16.

146. Mersch J, Jackson MA, Park M. et al. Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian. Cancer. 2015;121(2):269–275.

147. Jae Myoung Noh, Doo Ho Choi, Hyejin Baek. et al. Associations between BRCA Mutations in High-Risk Breast Cancer Patients and Familial Cancers Other than Breast or Ovary. J Breast Cancer. 2012;15(3):283–287.

148. Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V., and Boehning, D. (2012). Structure-Function of the Tumor Suppressor BRCA1. Comput. Struct. Biotechnol. J. 1, e201204005.

149. Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., et al. (2001). The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-Derived Mutation. J. Biol. Chem. 276 (18), 14537–14540.

150. Zhang, F., Fan, Q., Ren, K., and Andreassen, P. R. (2009). PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. Mol. Cancer Res. 7, 1110–1118.

151. Fu X, Tan W, Song Q, Pei H, Li J. BRCA1 and Breast Cancer: Molecular Mechanisms and Therapeutic Strategies. Front Cell Dev Biol. 2022 Mar 1;10:813457.

152. F. Wang, Q. Fang, Z. Ge, N. Yu, S. Xu, and X. Fan, "Common BRCA1 and BRCA2 mutations in breast cancer families: a meta-analysis from systematic review," Molecular Biology Reports, vol. 39, no. 3, pp. 2109–2118, 2012.

153. N. Tung, C. Battelli, B. Allen et al., "Frequency of mutations in individuals with breast cancer referred for BRCA1 and BRCA2 testing using next-generation sequencing with a 25-gene panel," Cancer, vol. 121, no. 1, pp. 25–33, 2015.

154. Y. Zhu, J. Wu, C. Zhang et al., "BRCA mutations and survival in breast cancer: an updated systematic review and meta-analysis," Oncotarget, vol. 7, pp. 70113–70127, 2016.

155. J. Zhang, R. Pei, Z. Pang et al., "Prevalence and characterization of BRCA1 and BRCA2 germline mutations in Chinese women with familial breast cancer," Breast Cancer Research and Treatment, vol. 132, no. 2, pp. 421–428, 2012.

156. Fernandes VC, Golubeva VA, Di Pietro G, Shields C, Amankwah K, Nepomuceno TC, *et al.* Impact of amino acid substitutions at secondary structures in the BRCT domains of the tumor suppressor BRCA1: Implications for clinical annotation. J Biol Chem. 2019 Apr 12;294(15):5980-5992.

157. Petitalot A, Dardillac E, Jacquet E, Nhiri N, Guirouilh-Barbat J, Julien P, Bouazzaoui I, Bonte D, Feunteun J, Schnell JA, Lafitte P, Aude JC, Noguès C, Rouleau E, Lidereau R, Lopez BS, Zinn-Justin S, Caputo SM; UNICANCER

Genetic Group BRCA network. Combining Homologous Recombination and Phosphopeptide-binding Data to Predict the Impact of BRCA1 BRCT Variants on Cancer Risk. Mol Cancer Res. 2019 Jan;17(1):54-69.

158. Findlay GM, Daza RM, Martin B, Zhang MD, Leith AP, Gasperini M, Janizek JD, Huang X, Starita LM, Shendure J. Accurate classification of BRCA1 variants with saturation genome editing. Nature. 2018 Oct;562(7726):217-222.

159. Parsons MT, Tudini E, Li H, Hahnen E, Wappenschmidt B, et al.. Large scale multifactorial likelihood quantitative analysis of BRCA1 and BRCA2 variants: An ENIGMA resource to support clinical variant classification. Hum Mutat. 2019 Sep;40(9):1557-1578.

160. Spina C., et al. Vitamin D and cancer. Anticancer Res. 2006;26(4A):2515–2524.

161. Jiang Y.J., et al. 1α, 25 (OH) 2-dihydroxyvitamin D3/VDR protects the skin from UVB-induced tumor formation by interacting with the β-catenin pathway. J. Steroid Biochem. Mol. Biol. 2013;136:229–232.

162. Feldman D, Krishnan AV, Swami S, Giovanucci E, Feldman BJ. The role of Vitamin D in reducing cancer risk and progression. Nat Rev. 2014;14:342–57.

163. Köstner K., et al. The relevance of vitamin D receptor (VDR) gene polymorphisms for cancer: a review of the literature. Anticancer Res. 2009;29(9):3511–3536.

164. Nam SE, Jin T, Park KS, Saindane M, Noh WC, Yoo YB, Park WS, Yun IJ. Vitamin D receptor (VDR) mRNA overexpression is associated with poor prognosis in breast carcinoma. Ann Surg Treat Res. 2022 Oct;103(4):183-194.

165. McCullough, M.L., Stevens, V.L., Diver, W.R. et al. Vitamin D pathway gene polymorphisms, diet, and risk of postmenopausal breast cancer: a nested case-control study. Breast Cancer Res 9, R9 (2007).

166. Akilzhanova A, Abilova Z, Sikhayeva N, Shtefanov I, Makishev A, Adylkhanov T, Rakhypbekov T, Zhumadilov Z, Momynaliev K. Vitamin D Receptor Gene Polymorphisms and Breast Cancer Risk in Kazakhstan. Cent Asian J Glob Health. 2014 Jan 24;2(Suppl):95.

167. Kazemi SM, Esmaieli-Bandboni A, Veisi Malekshahi Z, Shahbaz Sardood M, Hashemi M, Majidzadeh K, Kadkhodazadeh M, Esmaili R, Negahdari B. Vitamin D

receptor gene polymorphisms and risk of breast cancer in Iranian women. Ann Med Surg (Lond). 2021 Dec 4;73:103150.

168. Lurie G, Wilkens LR, Thompson PJ, Carney ME, Palmieri RT, Pharoah PD, Song H, Hogdall E, Kjaer SK, DiCioccio RA, McGuire V, Whittemore AS, Gayther SA, Gentry-Maharaj A, Menon U, Ramus SJ, Goodman MT; Ovarian Cancer Association Consortium. Vitamin D receptor rs2228570 polymorphism and invasive ovarian carcinoma risk: pooled analysis in five studies within the Ovarian Cancer Association Consortium. Int J Cancer. 2011 Feb 15;128(4):936-43.

169. Chen H, Zhu J. Vitamin D receptor rs2228570 polymorphism and susceptibility to ovarian cancer: An updated meta-analysis. J Obstet Gynaecol Res. 2018 Mar;44(3):556-565.

170. Uitterlinden AG, Fang Y, van Meurs JB, Pols HA, van Leeuwen JP. Genetics and biology of vitamin D receptor polymorphisms. Gene. 2004;338:143–156.

171. Terry SF (2015) Obama's Precision Medicine Initiative. Genet Test Mol Biomarkers. 19(3):113–114.

172. Mehrgou A, Akouchekian M (2016) The importance of BRCA1 and BRCA2 genes mutations in breast cancer development. Med J Islam Repub Iran. 30:369.

173. Carol AP, Vincent C (2014) Breast Cancer Survival Defined by the ER/PR/HER2 Subtypes and a Surrogate Classification according to Tumor Grade and Immunohistochemical Biomarkers. *J. Cancer Epidemiol.*

174. Onitilo AA, Engel JM, Greenlee RT et al. (2009) Breast cancer subtypes based on ER/PR and Her2 expression: comparison of clinicopathologic features and survival. Clin Med Res. 7(1-2):4–13.

175. Collignon J, Lousberg L, Schroeder H et al. (2016) Triple-negative breast cancer: treatment challenges and solutions. Breast Cancer (Dove Med Press). 20(8):93-107.

176. Zaharia M, Gómez H. (2013) Cáncer de mama triple negativo: una enfermedad de difícil diagnóstico y tratamiento (Triple negative breast cancer: a difficult disease to diagnose and treat). Rev Peru Med Exp Salud Publica. 30(4):649-56.

177. Freda L, John Z, Vanlalhruaii, et al. (2020) Whole exome sequencing identifies the novel putative gene variants related with type 2 diabetes in Mizo population, northeast India. Gene.

178. Andrew V, Zothan Z, Jeremy LP et al. (2020) Pediatric leukemia could be driven predominantly by non-synonymous variants in mitochondrial complex V in Mizo population from Northeast India. Mitochondrial DNA A DNA Mapp Seq Anal. 31(6):245-249.

179. Vanlalhruaii TM, Vanlalbiakdiki SC, Zothansanga et al. (2020) Analysis of variants in mitochondrial genome and their putative pathogenicity in tuberculosis patients from Mizoram, North east India. Mitochondrion. 54:21-25.

180. C. S. Chakrabarti, M. Roy, N. K. Sengupta, R. Lalthantluanga, P. P. Majumder. Genetic relationships among some tribal groups inhabiting the north-eastern, eastern and sub-Himalayan regions of India. Ann. Hum. Genet. (2002), 66, 361–368.

181. Biswabandhu Bankura, Bishnupriya Basak, Prajjval Pratap Singh, Albert Vanlalruata, Arindam Chatterjee, Sudakshina Ghosh, Rakesh Tamang, Manjil Hazarika, Gyaneshwer Chaubey, Madhusudan Das. Northeast India: Genetic Inconsistency across Ethnicity and Geography. bioRxiv 2024.01.09.574778.

***Bio-data***

ANDREW VANLALLAWMA
Department of Biotechnology
Mizoram University
Aizawl -796 004, Mizoram, India
E-mail: andrew13johnhelen@gmail.com
Mobile: +91 - 8794647992

**Summary**

I have completed my Bachelor's and Master's degree in Biotechnology from St. Edmunds College, Shillong and Mizoram University respectively. I have completed M. Phil programme in Department of Biotechnology, Mizoram University, under the guidance of Prof. N. Senthil Kumar where the title of the thesis was "Mutational screening of key genes in RTK/RAS signalling pathway associated with Pediatric Leukemia in Mizo population". I am working in the field of Human Disease Genomics with a focus on Cancer. Currently I am pursuing my Doctoral degree from the same lab and my thesis is "Whole Genome Analysis and its putative Clinical Impact Associated with Breast Cancer in Mizo Population", where I aim to find out the genomic markers contributing towards Adverse Drug Reaction (ADR) on Breast Cancer patients that took chemotherapy. Having an interest in Genomics, I also have been trying to elucidate the Mizo genealogy, tracing back to our cultural divergence and relating it to the inherited genomic mutations that have been passing down from our ancestors that could help towards better clinical management of human disease.

**Educational QualificationB**

| **Degree** | **Board** | **Division** | **Month and Year of Joining** | **Month and Year of Passing** |
|---|---|---|---|---|
| HSLC (Class 10) | Mizoram Board of School Education | I | March 2007 | April 2008 |
| HSSLC (Class 12) | Mizoram Board of School Education | I | March 2010 | April 2011 |
| B.Sc. Biotechnology | North Eastern Hill University | I | June 2011 | March 2014 |
| M.Sc. Biotechnology | Mizoram University | I | Aug 2014 | July 2016 |
| M.Phil. Course | Mizoram University | I | Aug 2017 | February 2018 |

| | | | | |
|---|---|---|---|---|
| Work | | | | |
| M.Phil. Awarded | Mizoram University | 0 | Aug 2017 | October 2019 |

**Awards:**

1. Cleared the Council of Scientific and Industrial Research, **CSIR – NET, Lectureship in the year 2015.**
2. **Best Oral presentation in international workshop** on "Introduction to basic and advanced biomedical approaches for enhancing QOL in aging societies" held at Biomedical Research Institute, AIST, Tsukuba Science City, Japan under DIACENTER – SUNRISE PROGRAM supported by AIST & JST (Govt. of Japan) and DBT (Govt. of India) from 14 – 21 Oct 2018.
3. **Oral presentation** titled, "Pharmacogenomics of Triple Negative Breast Cancer in Mizo Population", at **International Conference** on Recent Advances in Science and Technology organized by Mizoram University and NEAST
4. **Oral presentation** titled**,** "BRCA1, BRCA2, CHEK2 and PALB2 mutations in Triple Negative Breast Cancer patients, Mizoram"**, in National Seminar** on "Biotechnology for Sustainable Biosphere" organised by the Department of Biotechnology, Mizoram University on the 30$^{th}$ June and 1$^{st}$ July, 2023.

**Skills developed:**

DNA/RNA Isolation, Gel electrophoresis, PCR, Sanger Sequencing, Bioinformatics tools and methods. Whole Genome/Exome Sequencing and Whole Mitochondria Sequencing data analysis.

**Experience:**

| **Name of Post** | **Name of Project** | **From** | **Till** |
|---|---|---|---|
| JRF | DBT – Advanced Level State biotech Hub | 01.08.2016 | 31.01.2020 |
| Data Analyst | DBT - Genomelndia Project on Cataloguing the Genetic Variation in Indians | 01.02.2020 | Present |

**Conference attended:**

- **Poster presentation on International Conference on** "Biodiversity, Environment and Human Health: innovations and Emerging trends (BEHIET 2018), organized by School of Life Sciences, Mizoram University and Association of Biotechnology and Pharmacy (ABAP), India from 12 – 14 Nov 2018, on "***Pathogenicity of gene mutation in RTK/RAS signalling pathway contributing to leukemogenesis***"
- **Poster presentation in "42nd Annual IMA Pre-Conference Continuing Medical Education**" organised by The Academic Wing of IMA, MSB on 13 December 2018 at Aijal Club, Mizoram on "***Human disease genetics research in Mizoram***".
- **Oral presentation** in National Seminar on "Biotechnology for Sustainable Biosphere" organised by the Department of Biotechnology, Mizoram University on the 30th June and 1st July, 2023

**Workshop / Trainings attended:**

1. Participant in workshop on "Techniques in Molecular Biology", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 14 – 18 Oct 2015.
2. Participant in workshop on "Understanding basic principles in Human Molecular Genetics", organized by DBT – BIF, Dept. of Biotechnology, Mizoram University, from 07 – 11 Sept 2016.
3. Participant in International workshop on "Cancer epidemiology", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 29 – 30 Oct 2016.
4. Participant in workshop on "Hands on training on DNA barcoding and Phylogenetics", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 20 – 25 March 2017.
5. Participant in workshop on "Homology modelling and Molecular Docking", organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 18 – 22 April 2017.
6. Participant in Science Communication Workshop (*SciComm. 101*) held on 6 June, 2017 at Mizoram University.
7. Participant in International Workshop on "Molecular Entomology", organized Global

initiative for Academic Networks, from 19 – 28 June 2017.

8. Participate in workshop on "Antibiotic awareness and Infection Control Program" organized by DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 14 Nov. 2017.
9. Participate in 3rd NER, Research training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics – Driven Approaches" co-organized by National Institute of Biomedical Genomics, Kalyani and DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 19 - 24 Nov. 2017.
10. **Acted as a Resource Person in a workshop** on "Hands –on training on Molecular research technique" organized by DBT – Institutional Biotech Hub, GZRSC, funded by Rashtriya Uchchatar Shiksha Abhiyan (RUSA), GZRSC from 31 May – 6 June, 2018.
11. Participate in 3rd NER, ADVANCED Research training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics – Driven Approaches" co-organized by National Institute of Biomedical Genomics, Kalyani and DBT – State Biotech Hub, Dept. of Biotechnology, Mizoram University, from 23 July - 31 July, 2018.
12. Participant in CME course/workshop on "The concept and Application of genomics in clinical Medicine", jointly organized by Civil Hospital Aizawl and Mizoram University, conducted by CSIR – Institute of Genomics and Integrative Biology, New Delhi in 11 August 2018.
13. Participant in International workshop on "Introduction to basic and advanced biomedical approaches for enhancing QOL in aging societies" held at Biomedical Research Institute, AIST, Tsukuba Science City, Japan under DIACENTER – SUNRISE PROGRAM supported by AIST & JST (Govt. of Japan) and DBT (Govt. of India) from 14 – 21 Oct 2018.
14. Participate in 3rd NER Hands – on training workshop on "Gene Cloning, Protein Biochemistry, Structure Biology and Bioinformatics" organised by DBT – NER Biotechnology / Bioinformatics Training Centre, Advanced Centre for treatment, Research & Education in Cancer (ACTREC), Kharghar, Navi Mumbai. India from 4 – 15 Feb 2019.

15. Participate in webinar on “Winning the Game of Publishing”, organized by Wiley Research, in collaboration with DeLCON: DBT-Electronic Library Consortium on 28 October 2021.

**Publications:**

1. S Sahana, A Sivadas, M Mangla, A Jain, RC Bhoyar, K Pandhare, et al. Pharmacogenomic landscape of COVID-19 therapies from Indian population genomes. **Pharmacogenomics**. 2021. June. **IF = 2.8**
2. F Lalrohlui, J Zohmingthanga, V Hruaii, A Vanlallawma, NS Kumar. Whole exome sequencing identifies the novel putative gene variants related with type 2 diabetes in Mizo population, northeast India. **Gene**. 2021. 769, 145229. **IF = 3.5**
3. Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, et al. IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. **Nucleic acids research.** 2021.49 (D1), D1225-D1232. **IF = 14.9**
4. L Biakzuala, V Hrima, M Vanlalchhuana, A Vanlallawma, M Vabeiryureilai, et al. Contributions to Lycodon zawi, a little-known colubrid snake (Reptilia: Serpentes: Colubridae). **Herpetological Journal** 2020. 30 (4). **IF = 1.0**
5. A Vanlallawma, Z Zami, JL Pautu, Z Bawihtlung, L Khenglawt, et al. Pediatric leukemia could be driven predominantly by non-synonymous variants in mitochondrial complex V in Mizo population from Northeast India. **Mitochondrial DNA Part A.** 2020. 31 (6), 245-249. **IF = 1.69**
6. S Sivasubbu, V Scaria, et al. Genomics of rare genetic diseases—experiences from India. **Human genomics**. 2020. 13 (1), 1-18. **IF = 6.481.**
7. Jain, A., Bhoyar, R.C., Pandhare, K. et al. Genetic epidemiology of autoinflammatory disease variants in Indian population from 1029 whole genomes. **Journal of Genetic Engineering and Biotechnology.** (2021). 19, 183. **IF = 3.5.**
8. Vanlallawma, A., Lallawmzuali, D., Pautu, J.L. et al. Whole exome sequencing of pediatric leukemia reveals a novel InDel within FLT-3 gene in AML patient from Mizo tribal population, Northeast India. **BMC Genom Data.** 23, 23 (2022). **IF =2.9.**
9. E Zomawia, Z Zami, A Vanlallawma, NS Kumar, J Zothanzama, L Tlau, L Chhakchhuak, L Pachuau, JL Pautu, EVL Hmangaihzuali. Cancer awareness, diagnosis and treatment needs in Mizoram, India: evidence from 18 years trends (2003–2020). **The Lancet Regional Health – Southeast Asia**. 17 (2023)

## PARTICULARS OF THE CANDIDATE

| | | |
|---|---|---|
| NAME OF THE CANDIDATE | : | Andrew Vanlallawma |
| DEGREE | : | Ph.D. |
| DEPARTMENT | : | Biotechnology |
| TITLE OF THE THESIS | : | Whole Genome Analysis and its putative Clinical Impact Associated with Breast Cancer in Mizo Population. |
| DATE OF ADMISSION | : | 29.05.2020 |

APPROVAL OF RESEARCH PROPOSAL

| | | |
|---|---|---|
| 1. DRC | : | 20.10.2020 |
| 2. BOS | : | 27.10.2020 |
| 3. SCHOOL BOARD | : | 04.11.2020 |
| MZU REGISTRATION NO. | : | 1506957 |
| Ph.D. REGISTRATION NO. & DATE | : | MZU/Ph.D./1726 of 29.05.2020 |
| EXTENSION (IF ANY) | : | N/A |

**Head**

**Department of Biotechnology**

# ABSTRACT

# WHOLE GENOME ANALYSIS AND ITS PUTATIVE CLINICAL IMPACT ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

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

**ANDREW VANLALLAWMA**

**MZU REGISTRATION NO.: 1506957**

**Ph. D. REGISTRATION NO.: MZU/Ph. D./1726 of 29.05.2020**


**DEPARTMENT OF BIOTECHNOLOGY**

**SCHOOL OF LIFE SCIENCES**

**MARCH 2024**

# ABSTRACT

## WHOLE GENOME ANALYSIS AND ITS PUTATIVE CLINICAL IMPACT ASSOCIATED WITH BREAST CANCER IN MIZO POPULATION

**BY**

**ANDREW VANLALLAWMA**

**Department of Biotechnology**

**Supervisor: Prof. N. SENTHIL KUMAR**

**Submitted**

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

## 1. INTRODUCTION

The human body comprises around 36 trillion cells in males and 28 trillion cells in females (Hatton et al., 2023). Cells undergo growth, division, and replacement as needed, but malfunctions can lead to abnormal cell proliferation or resistance to programmed cell death. This may result in the formation of tumors, which can be cancerous (malignant) or non-cancerous (benign). Cancerous cells, characterized by their ability to invade nearby tissues and spread through metastasis, often form solid tumors. Examples of malignancies that don't form solid tumors include blood cancers like Leukemia (Hanahan and Weinberg, 2000).

Cancer poses a significant global health challenge, with millions of new cases diagnosed annually (Bray et al., 2018). It encompasses various diseases affecting diverse tissues or organs, each with unique characteristics and treatment approaches. According to Globocan (2020), global cancer incidence and mortality were 19,292,789 and 9,958,133, respectively, with 49.26% in Asia. Breast Cancer (BC) has the highest Age-Standardized Incidence Rate (ASIR) globally (47.8 per 100,000) and the second-highest Age-Standardized Mortality Rate (ASMR) (13.6 per 100,000) (Sung et al., 2020). In India, BC ranks first in both ASIR (28.5 per 100,000) and ASMR (13.3 per 100,000), representing 26.3% of new cancer cases among females (Sung et al., 2021).

The breast comprises glands, ducts, and fatty tissue above the upper ribs and chest muscles. Breast cancer (BC) can originate from various parts, each named accordingly. Lobules, producing breast milk, lead to lobular cancers when their cells become malignant. Ducts, connecting lobules to the nipple, carry breast milk, and ductal cancers are the most common. The areola, forming the nipple, houses ducts converging into larger vessels; malignancy here is Paget disease, a lesser-known BC type. Connective tissue (stroma) envelops ducts and lobules, providing support; stromal cell malignancy is called Phyllodes tumor, a rare BC type. Blood and lymph vessels in the breast may lead to angiosarcoma, a less common vessel lining-related breast cancer (Henry et al., 2020).

Most breast cancers (BC) fall under the category of carcinomas, tumors originating from epithelial cells lining organs and tissues. The predominant BC type is adenocarcinoma, starting in the cells of milk ducts or lobules (milk-producing glands). The specific cells turning malignant determine the BC type, indicating whether it has metastasized or remains local (American Cancer Society).

Breast cancer (BC) can be categorized into Ductal Carcinoma In-Situ (DCIS) and Invasive/Infiltrating BC, based on whether malignant cells are localized or have metastasized. DCIS involves malignant cells confined to ducts without spreading to surrounding tissues, also known as non-invasive or pre-invasive BC. DCIS can progress to invasive BC with metastatic potential. Infiltrating BC includes Infiltrating Ductal Carcinoma (IDC), where the duct lining turns cancerous and invades nearby tissues, and Infiltrating Lobular Carcinoma (ILC), originating from the lobule's milk-producing glands. IDC is more common, accounting for 8 out of 10 BC cases, compared to ILC, which represents 1 in 5 BC cases (Sharma et al., 2010).

Some invasive breast cancers have unique characteristics and progression patterns, influencing treatment and prognosis. Despite being less common, these types can pose a greater threat. One such form is Triple Negative Breast Cancer (TNBC), known for its aggressiveness and lack of Estrogen Receptors (ER-), Progesterone Receptors (PR-), and Human Epidermal Growth Factor Receptor 2 (HER2-negative).

Breast cancer (BC) encompasses various subtypes, including common ones like DCIS and IDC, special features in TNBC and Inflammatory Breast Cancer, and less common types such as Paget disease of the Breast Cancer and Phyllodes tumor. Studies using gene expression analysis, particularly microarray techniques, have categorized BC subtypes. The work initiated by Perou et al. (2000) was validated by different research groups, leading to the classification of BC into five distinct molecular groups. After cluster analysis, these were broadly grouped into Estrogen receptor-positive (ER-positive) and Estrogen receptor-negative (ER-negative) groups, each with its own subtypes (Sørlie et al., 2001; 2003; Zhao et al., 2004).

ER-positive subtypes in breast cancer can be classified based on the expression of estrogen receptor (ER) and genes associated with luminal epithelial cells. The

subclassification of luminal tumors is debated, but two main categories are recognized: luminal A and B tumors. This classification is determined by distinctive gene expression, with luminal B tumors exhibiting a higher proliferation cluster and may or may not express HER-2. ER-negative subtypes are categorized into three groups: HER-2-positive, Basal-like Tumors, and Breast-like Tumors. These classifications help in understanding the molecular diversity of breast cancer (Calza et al., 2006; Hu et al., 2006).

Basal-Like Tumors (BLT), lacking ER and HER-2 expression, exhibit distinct gene expression patterns with prognostic value. They pose challenges in treatment prediction due to a wide range of gene expressions. Microarray-based comparative genomic hybridization (array-CGH) reveals BLT's greater genomic instability, predicting DNA double-strand breaks in DNA repair mechanisms. BLT was initially categorized as ER-negative, but TNBC emerged as a distinct classification. TNBC, further subdivided into six types, is an evolving classification with varying subtypes identified by different studies. TNBC exhibits a complex genome with mutations in genes like TP53, BRCA1, and others, showing population-specific signatures. The complexity makes TNBC challenging to treat, relying heavily on chemotherapy.

The success or failure of cancer treatment depends on various factors, including an individual's genomic profile. Precision medicine tailors treatments to an individual's genetic makeup, optimizing efficacy while minimizing toxic effects. Pharmacogenetic variations, which directly impact drug processing and effects, play a crucial role in clinical outcomes. Well-established prognostic biomarkers for breast cancer, such as ER, PR, and HER-2, guide personalized treatment strategies. Genes like CYP2D6 and CYP19A1 are pharmacogenic markers influencing drug metabolism in breast cancer treatment.

Variations in genes like CYP2D6, CYP19A1, and others play a role in treatment outcomes, categorized as pharmacogenetic variants. Tamoxifen metabolism, influenced by CYP2D6 variants, leads to classifications like extensive, intermediate, or poor metabolizers. Similarly, aromatase inhibitor efficacy involves CYP19A1 variants. TNBC predisposing genes, including BRCA1, PALB2, FANCM, RAD51D,

ATM, and others, show population-specific mutations. The genomic complexity within TNBC poses challenges in treatment strategies, mainly relying on chemotherapy.

Population-based genomic studies in Mizoram, India, can reveal the genetic makeup, migration patterns, and polymorphisms' inheritance, offering valuable insights for disease implications and historical archives. India's genetic diversity, classified ethnically and linguistically, reflects migration patterns. However, comprehensive genomic studies are limited, and recent initiatives like IndiGenome provide valuable insights into population-specific variations.

Advancements in high-throughput sequencing platforms and decreasing costs have led to global sequencing projects, generating extensive datasets for population-specific variants. Mizoram, with a high incidence of cancer, necessitates scientific studies to understand the genomic variations, disease burden, and potential preventive measures. The study aims to identify pharmacogenomically important variants in TNBC patients in Mizoram, creating a reference for common polymorphisms in healthy individuals. This reference can be crucial for filtering disease-causing variants in large-scale NGS data analysis.

## 2. Objectives

1. Identification of genetic variants associated with pharmaco-genes in Triple Negative Breast Cancer.

2. Interpretation of population diversity using common polymorphic variants of Mizo tribal population.

## 3. Meterials and Methods

*Sample Collection*

In order to conduct this current study, Institutional Ethics Committee approval was obtained from Mizoram State Cancer Institute (MSCI)-No.D.12016/2/2013-MSCI/IEC (Appendix I). A total of 28 TNBC patients who are admitted in MSCI were recruited for the study after obtaining a signed consent form. The TNBC patients recruited for the study has a median age of 52 years with the oldest being 91 years and youngest being 24 years, providing a wider range for studying if age has an effect in the efficacy of chemotherapy drugs administered to them. Their clinical records were also collected using a well-structured questionnaire. Peripheral blood of 5 ml was drawn from each TNBC patients Ethylenediamine tetra-acetic acid (EDTA) coated vials and was stored in $-20^0$C until it was ready for DNA isolation.

Similarly, 20 individuals who are self-proclaimed healthy were taken as healthy controls with the possibility of using their genomic data to study the diversity with other populations. These 20 individuals were carefully selected based on: genetically unrelated, different sub-tribes representing the different districts in-order to maintain sample diversity. These 20 individuals are from a sub-set of a project entitled “Genomics for Public Health in India (IndiGenomes)” initiated by CSIR-Institute of Genomics and Integrative Biology, New Delhi.

*Adverse Drug Reaction Information*

To study the possible ADRs with different treatment regime, a well-structured questionnaire was formulated to follow the uniform Clinical Practice Guidelines (CPG). For this study we have studied different CPGs from different countries and opted to follow the National Comprehensive Cancer Network (NCCN). The structured questionnaire contains information such as personal details, date of diagnosis, histopathology, tumour location and grade, MBR grade, auxiliary lymph node involvement, types of chemotherapy drugs administered and no of cycles, TNM staging and different types of ADRs. The questionnaire was followed-up in 6 months interval to check their vital status and to record any complications they had endured..

*Univariate Binary Logistic Regression*

The questionnaires content such as clinical and histopathological records like tumour grade, MBR grade, auxiliary lymph node involvement, TNM staging, the combinations of chemotherapy administered and the number of cycles can affect the drug's efficacy. Similarly, WES analysis also reveals variants that are of pharmacogenomic importance, also there are variants that are also reported and classified as pathogenic in ClinVar database. In order to understand the relation of genomic variants, clinical and histopathological records, Spearman rank correlation co-efficient was used for significant level (p-value) <0.05 and confidence level of 95% were maintained using IBM Statistical Package for Social Sciences (SPSS) version 22.0.

*DNA Isolation and Visualization*

DNA isolation from TNBC patient blood samples for Whole Exome Sequencing (WES) was performed using the QIAamp® Blood Mini Kit. According to the manufacturer's protocol with few modifications. Agarose gel electrophoresis was performed to visualized the eluted genomic DNA.

*Whole Exome Sequencing and Data Analysis*

WES on 17 TNBC samples was carried out at the National Institute of Biomedical Genomics, Kalyani, India. The Whole Exome Analysis Pipeline (WEAP) was used for analysis, following GATK guidelines. BWA-MEM (Li 2013) was used to aligned the raw trimmed files which had already passed the quality checking, against the reference genome - GRCh38.p13 generating a SAM file which was converted to BAM file. The BAM file was then sorted using samtools utility (Li et al., 2009) according to their respective genomic coordinates. PCR duplicates were removed by using Picard tool as they can provide a false positive information and GATK Baserecalibrator utilizing a known variant dataset, "dbSNP138.vcf," was used to recalibrate base quality scores. To identify nucleotides that are not matching with the reference genome (called as variants) GATK HaplotypeCaller was used. This tool is used to identify variants present in germline as our sample was whole blood. To

accurately call germline variants GATK VariantRecalibrator was used to remove sequencing artifacts.

The variants that passed quality check is in a Variant Call Format (VCF) format. This file contains few details of the variants such as chromosome location, the variant position, the reference allele, the alternate allele and some others. To have a better understanding of a variant, annotation by ANNOVAR was performed. ANNOVAR utilises a number of publicly available databases such as RefGen, avsnp15, exac03, esp3500, ljb26, gnomad211exome, ClinVar, and 1000g. Such databases have their own niche of informations such as variant location, name of gene, function of gene, allele frequency, clinical impact, pathogenicity prediction etc

*Whole Genome Sequencing and Analysis*

Whole Genome Sequencing was performed at CSIR-Institute of Genomics and Integrative Biology, New Delhi as a collaborative effort. The libraries were prepared using TruSeq DNA PCR free library preparation kit as per manufacturer's instructions (Illumina Inc. San Diego, CA, USA, Cat. no. FC-121-9006DOC). Sequencing of the prepared libraries was performed on Illumina NovaSeq 6000 platform (Illumina Inc. San Diego, CA, USA). Alignment to reference genome, refining and default quality filtered variant calling was performed on the Illumina DRAGEN v3.4 Bio-IT platform (Illumina Inc. San Diego, CA, USA) usingGRCh38 as a human reference genome. Sentieon was used for joint variant calling that mimic the tools present in GATK and also calculates the allele frequency, allele count and allele number for the observed variants.

*Genes for Identifying Pharmacogenomic Variants*

WES annotated file contains approximately 30,000 variants per individual. This suggests that there are lots of variants that can be classified as variants of insignificance based on the impact towards contributing to disease manifestation, disease progression, susceptibility to, protective, risk factor and drug metabolism. To avoid such all variants, filtering steps were applied. The first filtering step used was "The Very Important Pharmacogenes" which is listed out in the PharmGKB database (later referred to as PharmGKB gene list). The list provides a well curated genes that

confers a significant role in the metabolism or response to one or a combination of multiple drugs. The list contains 69 genes while analysis on this study was being conducted.

In this study, majority of the TNBC patients received almost the same combination of chemotherapy drugs such as 5-Fluorocil, Epirubicin, Cyclophosphamide and Paclitaxel with Capecitabine, Doxorubicin, Docetaxel as a substitute in few of the patients. All the genes involved in transport and metabolizing these drugs were listed. Apart from this list of genes, genes that have pharmacogenomic importance that are listed in DrugBank was also utilized. DrugBank provides a list of genes that involves in the metabolism and transport of drugs. Finally, genes that were still absent in this list but have importance were obtained from literature survey which was later overlapped with DrugBank database. This list contains 88 genes

*Identification of Pharmacogenomic Variants*

From the 17 TNBC WES annotated file, filtering steps were done separately by matching the annotated file of TNBC patients with PharmGKB gene list and the same with metabolizing gene list. The second filter retained only those variants which are exonic and splice site in function. The next filtering retained only those variants that are predicted to be pathogenic/deleterious by any two of the pathogenicity predicting tools- Sorting Intolerant from Tolerant (SIFT), Polymorphism Phenotyping v2 (PolyPhen v2) and Mutation Taster. The resulting variants were checked using PharmGKB and ClinVar.

*Pathway Analysis Using DAVID*

Pathway analysis was performed using the Database for Annotation, Visualization, and Integrated Discovery (DAVID) on variant-associated genes to uncover functional annotations and interactions. A total of 111 genes were selected for pathway analysis. From the Annotation results further analysis was done using KEGG pathway and REACTOME pathway.

*Identification of Druggable Genes*

In this study, the TNBC annotated files were converted to .MAF files using perl script, this was then used in R-Package as an input file. The Maftools in R-package was used to analyse the MAF file to find the Drug-Interaction Function using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions.

*Inter-Population Differences in Allele Frequency*

Fisher Exact test compared allele frequencies of identified variants between TNBC patients and populations in databases like IndiGenomes, 1000 Genomes Project, ExAC, gnomAD, and esp6000.

*Establishing Common Polymorphic Variants Among TNBC Patients*

Common polymorphic variants were determined by excluding low-frequency, disease-associated, and non-benign variants. These variants were then matched with healthy controls and those that were also present in healthy controls were considered to be common polymorphic among mizo population.

*Interpretation of Population Diversity Using PCA Plot*

Principal Component Analysis (PCA) assessed genetic and geographical distances between populations based on SNP data from TNBC patients and the 1000 Genomes Project.

## 4. Results

A total of 28 breast cancer (BC) patients diagnosed with triple-negative breast cancer (TNBC) were recruited from Mizoram State Cancer Institute between 2018 and 2020. After obtaining signed consent forms, their medical records were collected, including information on diagnosis, treatment, and adverse drug reactions (ADRs). Clinical data such as histopathology, tumor grade, Modified Bloom-Richardson (MBR) grade, axillary lymph node involvement, chemotherapy treatments, and ADRs were obtained through a structured questionnaire.

Additionally, peripheral blood samples of 5 ml each were collected from the TNBC patients to investigate genomic alterations and their potential implications in cancer progression and drug toxicity.

A total of 30 ADRs were observed from 28 TNBC patients. The top 5 ADR observed are Alopecia (89.28%), Nausea (75%), Drowsiness (64.28), Change in appetite (53.57%) and Constipation (46.42%). The most severe ADR was Alopecia and majority of the of ADR severity observed was moderate.

Univariate Binary Logistic Regression was performed to reveal the association between the ADR and factors like mutations, clinical records such histopathology, grade of the tumour, axillary lymph node and the combinations of chemotherapy administered. Our analysis also revealed that patients who had Axillary Lymph Node involvement are likely to have Fatigue as an ADR with an OR of 8.889 (p-value = 0.023, 95% CI = 1.344 - 58.796). Analysis showed that patients having BRCA1.C1650R mutations are likely to have Gastritis with an OR of 22 (p-value = 0.031, 95% CI = 1.334 - 362.916). Patients with CYP4F2 mutations are also likely to Leukopenia with an OR of 13.06 (p-value = 0.034, 95% CI = 1.225 - 151.045). On the other hand, patients with the same mutation are less likely to have Drowsiness as an ADR with an OR of 0.074 (p-value = 0.034, 95% CI = 0.007 - 0.817).

After hard filtering steps WES analysis reveals a total of 125 variants were observed that could potentially play a role in clinical settings using PharmGKB database and DrugBank. We identify 4 variants that were having toxic effects with chemotherapeutic drugs and were present in PharmGKB database. DPYD (p.V732I, rs1801160) showed toxic effect to patients that were treated with fluorouracil with level of evidence score as 1A. MTHFR (p.A222V, rs1801133) showed toxic effect to patients that were treated with fluorouracil and capecitabine with level of evidence score as 4. SLCO1B1 (p.V174A, rs4149056) showed toxic effect to patients that were treated with cyclophosphamide; docetaxel; doxorubicin; epirubicin; fluorouracil with level of evidence score as 3. CYP2B6 (p.R22C, rs8192709) showed toxic effect to patients that were treated with cyclophosphamide with level of evidence score as 3.

We also observed 11variants that were showing pathogenic in three predicting tool and are also reported in ClinVar. rs5030737 (MBL2:p.R52C), rs80356993 (BRCA1:p.C1650R), rs56161402 (TPMT: p.R200H), rs56161402 (TPMT: p.R200H), rs34295337 (ERCC3:p.R109X), rs1801133 (MTHFR: p.A263V), rs4149056 (SLCO1B1: p.V174A), rs2228570 (VDR: p.M1?), rs186364861 (NUDT: p.V18I), rs1085307902 (BRCA1: p.Q79X), rs2108622 (CYP4F2: p.V433M), rs137852314 (G6PD: p.G163S ) were found in ClinVar. rs2228570 (VDR: p.M1?) was the most prevalent variant found in 16 samples, followed by rs2108622 (CYP4F2: p.V433M) found in 8 samples. rs80356993 (BRCA1:p.C1650R) was reported in many other studies and could be a diagnostic biomarker for the population as 3 patients harbours this mutation. However, validation by larger sample size will be required

We also observed twenty-four variants that had no rsID (i.e., not reported in dbSNP) and twenty-six variants that were reported in ClinVar with no clinical significance as well as its pathogenicity in disease conditions.

Using the observed important pharmacogenomic variants obtained in our study, we compare the allele frequency with different databases such as 1000 genomes, ExAc, ESP and gnomAD and found no association leading to a notion that it could be a population specific variant. KEGG pathway analysis revels top 14 pathways that were implicated with the genes analysed; including ABC transporters, Chemical carcinogens – DNA adducts, Chemical carcinogenesis – receptor activation, mineral absorption, MAPK signalling pathway. Reactome pathway analysis showed 45 possible pathways: including Drug Absorption Distribution Metabolism and Excretion (ADME), Transport of small molecules, Xenobiotics, CYP2F1 reactions, ABC-family proteins mediated transport, Metabolic disorder of biological oxidation enzymes. Drug-Interaction Function was identified using the Drug Gene Interaction database to reveal the druggable gene categories and drug–gene interactions. A total of 11 druggable categories and 11 druggable genes were identified.

Targeted panel sequencing of BRCA1, BRCA2, CHEK2 and PALB1reveals 39 variants with their frequency of prevalence among TNBC patients. Out of which 25

non-synonymous variants were in exonic region, 11 were synonymous, 3 non-synonymous in intronic region and 1 in 5'- UTR region.

The WES data of 17 TNBC patients were clubbed together and variants that has allele frequency lower than 0.5 as well as duplicates were excluded. Variants with no disease association in ClinVar were retained. Variants that had predicted to be benign in all three pathogenic score such as SIFT, Polyphen2 and Mutation Taster were then filtered out. These variants were then matched with healthy controls and variants that were also present in healthy controls were considered to be common polymorphic among Mizo population. A total of 35756 variants were obtained. Principal Component Analysis was performed Using R-package and reveals that using First principal component and Third principal component, Mizo population clustered with South Asian population.

## 5. DISCUSSION

In this study, 28 TNBC patients admitted to the Mizoram State Cancer Institute participated, providing a diverse sample cohort. The median age of the patients was 52, ranging from 24 to 91 years, allowing for an exploration of potential age-related effects on chemotherapy efficacy. Clinical records, including histopathology, tumor grade, Modified Bloom-Richardson (MBR) grade, axillary lymph node involvement, chemotherapy treatment details, and information on adverse drug reactions (ADRs), were collected through a comprehensive questionnaire.

Whole Exome Sequencing (WES) was conducted on 17 TNBC patients, and Whole Genome Sequencing (WGS) was performed on 20 healthy volunteers. The genomic data were analyzed to identify variants and mutations that might be associated with TNBC and its treatment outcomes.

The study focused on investigating the association between ADRs and various factors, including genomic mutations and clinical records. Univariate Binary Logistic Regression was employed for statistical analysis, aiming to reveal

correlations between ADRs and parameters such as histopathology, tumor grade, axillary lymph node involvement, and specific chemotherapy combinations administered. The study explored the role of BRCA1, a well-known susceptibility gene in familial breast cancer (BC) and ovarian cancer (OC), in TNBC. BRCA1 has been extensively studied for its involvement in DNA damage response and genomic stability, particularly in processes like Homologous Recombination (HR) and Non-Homologous End Joining (NHEJ).

The literature suggests BRCA1's relevance not only in BC and OC but also in other cancers such as prostate and pancreatic cancer. Emerging evidence has reported the presence of germline mutations in BRCA1 in gastric cancer (GC). The study aims to understand the implications of such mutations in the context of TNBC and its treatment outcomes.

The next steps involve further analysis of the genomic data, pathway analysis using tools like DAVID, identification of druggable genes, exploration of inter-population differences in allele frequencies of pharmacogenomic variants, and the establishment of common polymorphic variants among TNBC patients. Additionally, Principal Component Analysis (PCA) will be used to assess genetic and geographical distances between populations. The study holds promise for shedding light on the genetic factors influencing TNBC outcomes and treatment responses.

The association between BRCA1 mutations and gastric cancer (GC) has been studied extensively, particularly in the context of the Correa cascade, where prolonged gastric inflammation can lead to the progression from metaplasia to dysplasia and ultimately to carcinoma (Correa et al., 1992). Somatic loss of the BRCA1 allele has been observed in gastric malignancy, suggesting a potential role in predisposition to gastric cancer (Moiseyenko et al., 2013).

Studies have indicated a significant association between BRCA1 mutation carriers and gastritis in GC patients older than 50 years, compared to younger patients (Avanesyan et al., 2020). Moreover, first-degree relatives of BRCA1 mutation carriers have shown a 4-fold increase in the risk of developing GC, highlighting the potential role of BRCA1 in gastric cancer development (Buckley et

al., 2022). In the present study, patients with BRCA1.C1650R mutations were found to have a significantly higher likelihood of developing gastritis (OR = 22, p-value = 0.031, 95% CI = 1.334 - 362.916).

Additionally, patients with BRCA1.C1650R mutations were also observed to be more likely to experience leukopenia (OR = 13.06, p-value = 0.034, 95% CI = 1.225 - 151.045). Among the TNBC patients in the study, 32.14% were known to have leukopenia, with varying degrees of severity. Notably, a retrospective matched cohort study in the United States reported that leukopenia was the most severe adverse drug reaction (ADR) observed in BRCA1 carriers treated with cyclophosphamide, taxanes, and carboplatin-based therapy (West et al., 2019). Although the current study population did not receive carboplatin, the observed leukopenia aligns with the broader literature on BRCA1-related severe ADRs during chemotherapy, particularly with anthracycline and cyclophosphamide-based regimens (Furlanetto et al., 2021). Despite the significance of these findings, the study acknowledges the limitation of a small sample size, emphasizing the need for further investigations with larger cohorts.

The analysis revealed several significant associations between clinical parameters and adverse drug reactions (ADRs) among TNBC patients. Specifically, patients with axillary lymph node involvement were found to be more likely to experience fatigue as an ADR (OR = 8.889, p-value = 0.023, 95% CI = 1.344 - 58.796). This aligns with previous studies indicating fatigue as a common side effect of neo-adjuvant chemotherapy, particularly with anthracycline, cyclophosphamide, and taxane-based regimens (Kumar et al., 2019). The presence of axillary lymph node involvement is an important prognostic factor in TNBC patients; however, its correlation with chemotherapy-related ADRs underscores the complexity of treatment considerations. While axillary lymph node involvement is a prognostic indicator, treatment decisions should also consider other clinicopathological factors to optimize disease-free and overall survival rates within the population (Houvenaeghel et al., 2016).

Additionally, the analysis identified a total of 125 variants from the WES data of TNBC patients after variant prioritization and filtering steps. These variants were categorized based on their presence in pharmacogenomic databases, such as PharmGKB and ClinVar, as well as their implications in other diseases or lack of rsID. Among the identified variants, 24 were found to have no rsID, potentially indicating population-specific variants warranting further investigation. Notably, variants in genes like MTHFR (rs1801133), SLCO1B1 (rs4149056), and DPYD (rs1801160) were observed in both PharmGKB and ClinVar databases, suggesting their relevance in pharmacogenomic considerations for TNBC treatment.

The smaller number of variants obtained post-filtering reflects the stringent criteria applied to prioritize variants with known pharmacogenomic relevance. While a large number of variants were initially observed, the focus on curated pharmacogenomic databases ensured the selection of variants most likely to contribute to disease manifestation and drug response. However, further exploration of variants without rsID and those with potential clinical significance in other diseases is warranted to comprehensively understand their role in TNBC pathogenesis and treatment response.

*MTHFR*

*MTHFR* is responsible in folic acid metabolism, that converts 5,10-methylenetetrahydrofolate into 5-methyltetrahydrofolate, the main form of folate essential in various cellular processes. This includes the synthesis of thymidine in DNA, crucial for DNA synthesis and repair. Mutations in the MTHFR gene can result in reduced enzyme activity, leading to the accumulation of 5,10-methylenetetrahydrofolate and impacting the remethylation process, ultimately causing elevated homocysteine levels in the blood. Increased homocysteine levels have been associated with several cancers, including breast cancer (BC), colorectal cancer, prostate cancer, and leukemia.

The most common MTHFR gene mutations linked to breast cancer risk are p.C677T and p.A1298C variants. The p.C677T variant, particularly the CT and TT genotypes, has been associated with an increased risk of breast cancer in various

non-Asian populations, such as Jordanian, Iranian, and Latin American populations, as well as in mixed populations including Asian and Caucasian. However, there are contradictory findings in other populations, such as the Brazilian and North American populations, suggesting population-specific mutational signatures.

Our study identified a novel variant (G>A, c.C788T, p.A222V) in the MTHFR gene associated with toxicity towards capecitabine and fluorouracil in 23.52% of TNBC patients. The PharmGKB database reports that the AA genotype has the highest risk of toxicity, followed by the GA genotype, indicating an increased risk of toxicity when treated with capecitabine and fluorouracil compared to the wild-type genotype. All TNBC patients in our study received 5-fluorouracil, a thymidylate synthase inhibitor. The identified MTHFR variant may result in a loss of function or reduced enzyme activity, leading to an accumulation of 5,10-methylenetetrahydrofolate and subsequently higher homocysteine levels in the bloodstream, potentially influencing cancer development.

It is crucial to note that the functional implications of this variant are not yet fully understood, and further experiments are necessary to establish its role in toxicity against capecitabine and fluorouracil. Additionally, the allele frequency of the GA genotype needs to be studied in a larger cohort to determine whether the Mizo population is predisposed to this variant. This underscores the importance of population-specific investigations to understand the unique genetic factors contributing to cancer susceptibility and drug toxicity.

*SLCO1B1*

SLCO1B1, part of the Solute Carrier (SLC) superfamily, encodes organic anion-transporting polypeptides (OATPs) found in various organ membranes. OATPs play a crucial role in the absorption, distribution, and elimination of essential nutrients, toxins, and therapeutic drugs. Genetic variations in SLCO genes have been associated with alterations in the pharmacokinetics of their substrates, impacting drug efficacy and potential toxicities.

One prominent single-nucleotide polymorphism (SNP) in SLCO1B1, rs4149056 (T>C, p.V174A), has been linked to reduced transporting efficacy of

OATP1B1. This SNP results in significantly elevated plasma levels of various statins, affecting their therapeutic outcomes. Additionally, certain variants of SLCO1B1 have been associated with accelerated methotrexate clearance, potentially increasing the risk of gastrointestinal toxicity during methotrexate therapy, particularly in pediatric acute lymphoblastic leukemia.

In our study, although none of the patients received methotrexate or statins, the investigation of SLCO1B1 variant rs4149056 (T>C, p.V174A) in younger female breast cancer patients (age < 45) revealed its potential involvement in the development of chemotherapy-induced amenorrhea. This variant was identified as a predictor of chemotherapy-induced amenorrhea in these patients.

It's worth noting that the PharmGKB database has cataloged this SLCO1B1 variant to have a toxic effect when breast cancer patients are treated with cyclophosphamide, with a level of evidence 3. This underscores the importance of considering genetic variations in drug transporter genes, such as SLCO1B1, in understanding the variability in drug response and potential adverse effects in cancer patients undergoing chemotherapy. Further research and validation are essential to establish the clinical significance of these findings and to guide personalized treatment approaches.

*DPYD*

DPYD, encoding Dihydropyrimidine dehydrogenase (DPD), plays a crucial role in the catabolic process responsible for the clearance of 5-fluorouracil (5-FU). 5-FU undergoes enzymatic reactions within the cellular environment, leading to the generation of active metabolites. One such metabolite, fluorodeoxyuridine monophosphate, acts as a competitive inhibitor of thymidylate synthase, disrupting nucleotide availability crucial for cancer cell proliferation and inducing apoptosis. Additionally, 5-FU has the ability to integrate into DNA, initiating DNA damage repair mechanisms and apoptotic pathways.

Several DPYD variants have been identified as clinically relevant in the context of 5-FU toxicity. Notable variants include rs3918290, rs56038477, rs67376798, and rs55886062, which have been associated with 5-FU toxicity. A

comprehensive study among the South Asian population highlighted fifteen DPYD variants of importance related to 5-FU toxicity, emphasizing the need for further clinical investigations. In our study, the variant p.V732I (rs1801160) was identified, and it was noted to have the highest allele frequency among the studied cohorts, particularly in Asians. However, there was no significant association between this variant and toxic effects from 5-FU, consistent with findings from other studies, including those in the Chinese population.

The PharmGKB database also supports the lack of association between rs1801160 and 5-FU toxicity. Despite the amino acid substitution (p.V732I), the variant is assigned a normal function by the Clinical Pharmacogenetics Implementation Consortium (CPIC). This suggests that, based on current knowledge, this DPYD variant is not linked to increased susceptibility to 5-FU toxicity.

It's essential to consider individual genetic variations, like those in DPYD, when administering 5-FU-based chemotherapy to ensure optimal treatment outcomes and minimize the risk of adverse effects. Ongoing research and comprehensive understanding of these genetic factors will contribute to more personalized and effective cancer treatment strategies.

*BRCA1*

BRCA1, initially linked to susceptibility to breast and ovarian cancer, has been a pivotal gene in understanding familial cancer risk. Germline mutations in BRCA1 have been associated with heightened risks for various cancers, extending beyond breast and ovarian, to include prostate, colorectal, stomach, and pancreatic cancers. Approximately 20% of familial cases of breast and ovarian cancers are attributed to BRCA1 mutations. The gene's involvement in DNA repair mechanisms, particularly through its interactions with various partners like BARD1, PALB2, BRCA2, and others, highlights its crucial role in maintaining genome stability through homologous recombination.

Common germline mutations in the BRCA1 gene include 5382insC, 185delAG, 3819del5, and 4153delA, with 5% of breast cancer patients in Western populations potentially harboring heritable mutations. Notably, the prevalence of

BRCA1/2 mutations is generally lower in Asian populations compared to Caucasians. A specific missense mutation, c.5089T>C variant in BRCA1, leads to the substitution of Cysteine with Arginine at amino acid position 1697 (p.Cys1697Arg). This variant, found in gnomAD v3.1, falls within a crucial functional domain and has been reported as pathogenic in various studies.

Computational predictors, such as BayesDel, suggest a potential impact on BRCA1 function via protein alteration, and the SpliceAI predictor indicates no involvement in splicing. The saturation genome editing (SGE) assay also classifies the variant as loss of function. Interestingly, this variant is absent in gnomAD v3.1, and its potential predisposition to the Asian population makes it a candidate for a diagnostic biomarker. However, further validation with a larger number of samples is essential to establish its significance as a biomarker for diagnostic purposes. Continued research on specific BRCA1 mutations will contribute to understanding their implications, allowing for more targeted and personalized approaches in cancer diagnostics and treatment.

*Inter-population differences in the allele frequency of important pharmacogenomic variants*

Understanding inter-population differences in the allele frequency of important pharmacogenomic variants is crucial for advancing precision medicine, particularly in diseases like breast cancer. Precision medicine emphasizes individualized approaches based on the variability in genes, environment, and lifestyle. The identification of genomic changes associated with disease manifestations and progression allows for tailored treatments with fewer side effects.

Breast cancer, being extensively studied globally, has seen significant progress in biomarker development for early detection and precise treatment. However, certain subtypes of breast cancer pose challenges, requiring further exploration for improved early diagnosis and treatment strategies.

Comparing the allele frequency of key pharmacogenomic variants using the Fisher exact test revealed a significant association with the ExAc and gnomAD databases. ExAc, initially a stand-alone database focusing on exome sequencing data,

later merged with gnomAD, which encompasses both genomic and structural variant data. This association underscores the importance of these variants in both exonic and genomic contexts.

However, variations in the allele frequency across other databases indicate a mixed distribution. The observed differences suggest that these variants may be population-specific, warranting further investigation with larger sample sizes. As precision medicine continues to evolve, unraveling the genetic intricacies associated with pharmacogenomic variants will contribute to more effective and personalized disease treatments, ultimately improving clinical outcomes.

*Pathway analysis*

Pathway analysis was performed in DAVID, KEGG and Reactome databases were used for pathway enrichment. The genes identified from the WES data were subjected to pathway analysis using DAVID with the latest Knowledgebase v2023q2. KEGG pathway analysis highlighted the top 14 pathways associated with the analyzed genes, including ABC transporters, chemical carcinogens – DNA adducts, chemical carcinogenesis – receptor activation, mineral absorption, and MAPK signaling pathway. Reactome analysis identified 45 possible pathways, such as Drug Absorption Distribution Metabolism and Excretion (ADME), Transport of small molecules, Xenobiotics, CYP2F1 reactions, ABC-family proteins mediated transport, and Metabolic disorder of biological oxidation enzymes. Many of these pathways are related to cancer, indicating the involvement of genes in inducing carcinogenesis.

*Cataloguing of common polymorphic variants among Mizo population*

The large amount of data generated by next-generation sequencing (NGS) poses a challenge in identifying significant variants. NGS data covers approximately 60 million base pairs, generating around 30,000 to 40,000 variants per sample. Bioinformatic tools are crucial for eliminating variants that may be simple polymorphisms. However, certain variants may be specific to populations practicing endogamy, such as the Mizo population. Identifying common polymorphic variants in the Mizo population is essential for distinguishing between disease-associated variants and those that are population-specific and benign.

Understanding the genomic background of populations is crucial for providing targeted and personalized treatments. Next-generation sequencing methods have significantly advanced this goal by identifying DNA alterations and the associated genes. The genetic data available for the Mizo population, including exonic regions and mitochondrial DNA, highlight the importance of mapping, calling variants, annotating them, and interpreting their functional and clinical impacts. This knowledge is essential for tailoring healthcare interventions to the specific requirements of diverse populations.

*Cataloguing of common polymorphic variants among Mizo population*

Next-generation sequencing (NGS) data encompasses vast information, covering around 60 million base pairs, constituting about 2% of the genome. This data includes variants of significant value in disease association, as well as variants that may not play a role in diseases. Typically, an NGS experimental design with a coverage of 100x and 150 pair-end reads generates approximately 10 gigabytes of data per sample run, translating to about 30,000 to 40,000 variants per sample. The challenge lies in navigating through this extensive data to identify important and significant variants that may be key players in the studied disease.

Numerous bioinformatic tools are available to sift through variants, differentiating between those that may be simple polymorphisms for the population and those that are more prevalent in specific populations. These population-specific variants could be predisposed variants, conferring risks for certain diseases, exhibiting toxicity towards administered drugs, or even serving as causal factors for the population. The concept of predisposition has a pronounced impact on populations practicing endogamy, and the Mizo population, with a history of endogamy, exemplifies this effect.

For populations like the Mizo, it becomes crucial to identify variants that are common polymorphisms, distinguishing them from disease-associated variants. The identified variations may not fully capture the true diversity within the Indian population. In-house genetic data for the Mizo population, focusing on the exonic region and mitochondrial DNA, has been produced, indicating challenges in

mapping, calling variants, annotating them, and interpreting their functional and clinical impacts.

Understanding the genomic background of a population is essential for providing targeted treatments tailored to their specific needs. Advances in NGS methods offer a platform where technology can identify DNA alterations, including the genes where variants are present. Recognizing differences in the genomic background among populations is an urgent need for advancing personalized medicine and improving healthcare outcomes.

*PCA Plot*

In general, maximum variants were captured in first three principal components. Using R with principal component 1 and 2 plotted graph, the populations were color-coded (African – yellow, Caucasian – green, Hispanic – light grey, East Asian – blue, South Asian – black, Mizoram - red). Principal Component Analysis was performed Using R-package and reveals that using First principal component and Third principal component, Mizo population clustered with South Asian population