GENETIC VARIANTS IN PHARMACOGENES IN NORTH-EAST INDIAN POPULATION AND THEIR ASSOCIATION WITH ADVERSE DRUG REACTION IN GASTRIC CANCER

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

RANJAN JYOTI SARMA

MZU REGISTRATION NO.: 1703911

Ph. D. REGISTRATION NO.: MZU/Ph. D./1725 of 29.05.2020

DEPARTMENT OF BIOTECHNOLOGY

SCHOOL OF LIFE SCIENCES

OCTOBER 2023

GENETIC VARIANTS IN PHARMACOGENES IN NORTH-EAST INDIAN POPULATION AND THEIR ASSOCIATION WITH ADVERSE DRUG REACTION IN GASTRIC CANCER

BY

RANJAN JYOTI SARMA

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 **"Genetic Variants in Pharmacogenes in North-East Indian population and their association with Adverse Drug Reaction in Gastric Cancer"** submitted by **Ranjan Jyoti Sarma**, Ph.D. Research 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: October **Year: 2023**

DECLARATION

I **Ranjan Jyoti Sarma**, 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.**

(RANJAN JYOTI SARMA)
Research Scholar

(PROF. JOHN ZOTHANZAMA)
Head of Department

(PROF. N. SENTHIL KUMAR)
Supervisor

ACKNOWLEDGEMENT

I consider myself incredibly fortunate to have Prof N. Senthil Kumar, an esteemed Professor in the Department of Biotechnology at Mizoram University, as my supervisor. I am profoundly grateful to him for his unwavering support, invaluable guidance, scholarly expertise, continuous encouragement, and heartfelt well-wishes throughout my research journey. Without his presence, this work would never have reached completion. Prof. N. S. Kumar has not only provided moral and spiritual support but has also been a pillar of strength in every aspect of my research endeavour.

In addition, I would like to extend my sincerest appreciation to Prof. John Zothanzama, the Head of the Department of Biotechnology, and Dr. Th. Robert Singh, former Head of the Department of Biotechnology, Mizoram University. I am equally indebted to the esteemed faculty members of the department, including Dr. H. Lalhruaitluanga, and Dr. Satish Kumar Pandey, Assistant Professor of the Department of Biotechnology, Mizoram University for their unwavering support and encouragement during my research work. I am also grateful to the non-teaching staff of the Department of Biotechnology, Mizoram University.

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, Dr. Saia Chenkual, Zoram Medical College, Mizoram and other clinicians and surgeons from Civil Hospital, Aizawl and Mizoram State Cancer Institute, Mizoram.

I am immensely grateful to my fellow lab mates, Andrew Vanlallawma, Zothanzami, Lalfakzuali, Hlua Pachuau 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. Without his unwavering backing, I would not have achieved this significant milestone. Additionally, I would like to express my

gratitude to Dr. Manashjyoti Deka, Mr. Angshuman Das Tariang, and Mr. Nghaktea Fanai for their invaluable assistance during my stay at Mizoram University.

I extend my sincere appreciation to my beloved parents, Mr. Pradip Kumar Sarma and Mrs. Minati Devi, as well as my elder brother, Mr. Kishor Jyoti Sarma, sister-in-law, Mrs. Udiptya Bhattacharya, and elder sisters Mrs. Barnali Devi, Mrs. Meenakshi Sarma and Mrs. Purabi Sarma. Their boundless love, prayers, care, and moral support have been instrumental in my research accomplishments. I am profoundly grateful to my fiancée, Adv. Nilakshi Sarma, whose encouragement, and understanding sustained me through the challenges of this academic endeavour.

I would like to thank Dr. Payel Chakraborty and Dr. Souvik Ghatak, Post-Doctoral Fellow, Lund University, Sweden, Dr. Vineet Mishra, DBT RA, CSIR-IICB, Kolkata, and Dr. Ajit Kumar Passari, former scientist, SRUC Edinburgh Campus, United Kingdome, Dr. S. Sarathbabu, Mizoram University for their valuable suggestions in the study. I would like to thank Dr. Samrat Adhikari, Head, Department of Biotechnology, St. Edmund's College, Shillong and Dr. Ayaluru Murali, Assistant Professor, Department of Bioinformatics, Pondicherry University for inspiring me throughout my Ph.D. journey.

I am thankful to the Data Collectors and Lab Technicians, Jonathan Lalramhluna, Baby Lalrintluangi, R. Lalengkimi, and T. Lalhriatpuii 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.

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. Within the limited space of this acknowledgement, I strive to convey my deep appreciation to all those who have contributed to my research journey.

Place: Aizawl **(RANJAN JYOTI SARMA)**

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

5FU	:	5-Fluorouracil
ADME	:	Absorption, Distribution, Metabolism, And Excretion
ADR	:	Adverse Drug Reaction
AERS	:	Adverse Event Reporting System
ALK	:	Anaplastic Lymphoma Kinase
ALL	:	Acute Lymphoblastic Leukaemia
BAM	:	Binary Alignment/Map
bp	:	Base-pair
BQSR	:	Base Quality Score Recalibration
BWA	:	Burrows Wheeler Algorithm
COSMIC	:	Somatic Mutations in Cancer
CPIC	:	Pharmacogenetics Implementation Consortium
CTCAE	:	Common Terminology Criteria for Adverse Events
CTEP	:	Cancer Therapy Evaluation Program
CYP	:	Cytochrome P450
DAR	:	Dermatological Adverse Reaction
DPWG	:	Dutch Pharmacogenetics Working Group
DPYD	:	Dihydropyridine dehydrogenase
DPYS	:	Dihydropyrimidinease
DRAGEN	:	Dynamic Read Analysis for GENomics
ECG	:	Echocardiography
eMERGE	:	Medical Records and Genomics
EML4	:	EMAP Like 4
eQTL	:	Expression Quantitative Trait Loci
FLO	:	5-Fluorouracil , Leucovorin, Oxaliplatin
GATK	:	Genome Analysis Toolkit
GC	:	Gastric Cancer
GD	:	Gaucher disease
GEO	:	Gene Expression Omnibus
GRCh	:	Genome Reference Consortium Human

GTEx	:	Genotype-Tissue Expression
Hb	:	Hemoglobin
HCV	:	Hepatitis C Virus
HGNC	:	Hugo Gene Nomenclature System
HGP	:	Human Genome Project
HTML5	:	HyperText Markup Language 5
IEC	:	Institutional Ethics Committee
IndiGenomes	:	Genomics for Public Health in India
KM-Plotter	:	Kaplan Meier plotter
LoA	:	Loss of Appettite
MHC	:	Major Histocompatibility Complex
MPVardb	:	Mizoram Pharmacogenomic Variant Database
NES	:	Net Effect Size
NHGRI	:	Human Genome Research Institute
NSCLC	:	Non-small cell lung cancer
PGRN	:	Pharmacogenomics Research Network
PGx-Genes	:	Pharmacogene
PharmGKB	:	Pharmacogenomics Knowledgebase
PHP	:	Hypertecxt Preprocessor
PK/PD	:	Pharmacokinetics/ Pharmacodynamics
QTQ	:	QC-Trim-QC
SAM	:	Sequence Alignment/Map
SJS	:	Steven-Johnson syndrome
SNP	:	Single Nucleotide Polymorphism
SOB	:	Shortness of Breath
STAD	:	Stomach Adenocarcinoma
T2T- CHM13	:	Telomere-to-Telomere-CHM13
TAR	:	Thrombocytopenia with Absent Radii
TCGA	:	The Cancer Genome Atlas
TCGA	:	The Cancer Genome Atlas
TNM	:	Tumor, Node, Metastasis

TPMT2	:	Thiopurine methyltransferase
UDP	:	Uridine Diphosphate
UPB1	:	Beta-ureidopropionase
VDR	:	Variable Drug Responses
VIP	:	Very Important Pharmacogene
VQSR	:	Variant Quality Score Recalibration
WEAP	:	Whole Exome Analysis Pipeline
WES	:	Whole Exome Sequencing
WGS	:	Whole Genome Sequencing
WHO	:	World Health Organization

CHAPTER 1

INTRODUCTION AND REVIEW OF LITERATURE

1.1 Introduction

The human haploid whole genome is more than 3 billion base-pair (bp) long with an interindividual similarity of approximately 99.5%. However, due to population admixture, individuals from different populations are likely to be more similar than those from the same population (Witherspoon et al., 2007; Karki et al., 2015). The human reference genome provides centralized genomic coordinates that are useful for comparing genomic loci identified in other study results. Evolving technology in DNA sequencing played an unprecedented role in understanding genome diversity in terms of polymorphism in the genes in coding and non-coding regions (Wang et al., 2022). This has enabled to study of the implications of genetic polymorphism in common diseases such as cancers and rare diseases.

Common diseases can be polygenic and multiple-risk alleles, which may have cumulative effects along with epigenetic factors. However, rare diseases or Mendelian diseases (Huntington's disease, Lynch syndrome, familial CAD, thrombocytopenia with absent radii (TAR) syndrome, Gaucher disease (GD), etc.) are monogenic and often due to a single mutation (single nucleotide substitution) or deletion of a few base pairs from the important coding part or in the regulatory site of the non-coding regions of a gene. Moreover, different monogenic disorders occurring simultaneously may cause different phenotypes due to combinatorial effects (Zhang et al., 2015; Tahsin et al., 2020).

Common diseases such as different types of cancer are influenced by more than one genetic mutation. There are several germline gene mutations identified using polygenic risk scores derived from genome-wide association studies in prostate cancer, gastric cancer, breast cancer, and other common forms of cancers (Dian et al, 2021; Lopes Cardozo et al., 2023; Houlahan et al., 2023). Genetic polymorphism not only causes the disease state complex but also affects the prognosis of the disease by various means. Certain variations such as secondary mutation may develop resistance against a particular therapy when the therapy is provided to the patient for a long time. For

example, T790M, C797S and L792F mutation in *EGFR* develop resistance against tyrosine kinase inhibitors used in NSCLC, and L1196M, and G1269A in ALK develop resistance against *EML4-ALK* targeting drug by reducing the efficacy of the drug (Ma et al., 2011; Hamid et al., 2020).

Disease prognosis also gets affected by the variations in the gene involved in the process of drug metabolism. Genes involved in drug metabolism are referred to as pharmacogenes or PGx-Genes and are also responsible for variable drug responses (VDR). VDR could be influenced by the common as well as rare polymorphisms across a population and could be studied for pharmacogenomics and pharmacogenetics (Lesko et al., 2004). VDR is also one of the major causes of adverse drug reaction (ADR) due to any variable responses during the pharmacokinetic and pharmacodynamic pathways of the drug. Pharmacokinetics is defined by the body's overall actions on the drug which involves the absorption followed by distribution, metabolism, and excretion (ADME) of the drug molecule.

Drug metabolism occurs in multiple phases namely conjugation (Phase 1 involves Cytochrome P450), and glucuronidation (Phase 2 involves uridine diphosphate (UDP)-glucuronosyltransferases, sulfotransferases, and glutathione S-transferases), Sulphanation and acetylation and primarily occurs in the liver. It has been identified that slow aceylators characterized by a higher concentration of parent drug in urine tend to show more side effects than fast acetylators, characterized by a lower concentration of parent drug in urine (Shenfield et al., 2004). Biotransformation of the drug's lipophilic center to the hydrophilic center allows the drug to excrete via. bile or urine. Cytochrome P450 also attributed as *CYP*, a multi-gene family plays various important roles in drug metabolism and clearance of drugs and xenobiotics. Interethnic and interindividual variations in the therapeutic effectiveness of drugs pose a significant concern in healthcare, with genetic polymorphisms and epigenetic modifications in CYP genes, along with environmental factors, being potential contributors to these variations (Zhao et al., 2021).

Genes encoding CYP enzymes also referred to as drug-metabolizing enzymes are highly polymorphic and their allele frequency is considerably different between

the different populations. In contrast to pharmacogenetics which identifies SNPs in DNA sequences that cause clinically significant alterations of activity in drug-metabolizing enzymes, the population-level pharmacogenomic studies revealed several polymorphisms integrating whole genome or whole exome with their allele frequency distribution in a population that has impacts on drug metabolism (Ma et al., 2002, Tornio et al., 2018).

SNPs in the CYP family genes may have impacted the pharmacokinetics of commercial drugs. Polymorphism of CYP enzymes encoding genes has been a primary focus in pharmacogenetics since these enzymes are responsible for the metabolism wide range of marketed drugs. Pharmacogenetic testing has placed specific emphasis on genes such as *CYP2D6, CYP2C19, CYP2C9*, and *CYP3A4/5*, given that they encode the most prevalent and genetically diverse cytochrome P450 (CYP) enzymes crucial for drug metabolism (Finkelstein et al., 2016). The phase I metabolism of drugs might be greatly impacted by highly variable variations in *CYP2D6*, *CYP2C19*, and *CYP2C9* since these enzymes metabolize almost 70-80% of drugs in use today. It was found that less than 14% of Asians, Africans, and Caucasians also experienced *CYP2D6* deficiency, due to polymorphisms, and are classified as poor metabolizers. *CYP2C9* can be considered as the most clinically significant metabolizer as the multiple SNPs identified in the gene directly impact the efficacy of the drugs and are also responsible for ADR events. Moreover, several polymorphisms in CYP family genes are also involved in carcinogen bioactivation (Zhou et al., 2009; Preissner et al., 2013).

Significant polymorphic cytochrome P450 enzymes include 1A2, 2D6, 2C9, and 2C19, highlighting their importance in understanding genetic variations impacting drug metabolism (Preissner et al., 2013). *CYP3A4* is the most abundant hepatic enzyme (15 to 20%) involved in drug metabolism and activity is highly variable. An intronic variant *CYP3A4*22* (dbSNP: rs35599367) variant occurs in Europeans with minor allele frequency (MAF) of 5%, and admixed Americans with MAF 3%, but is rare in other populations with MAF < 0.01 (Tirona et al., 2017; Tornio et al., 2018). Another clinically important gene is *CYP2D6*, which metabolizes a wide range of antipsychotics and antidepressant drugs. *CYP2D6* genotyping is extremely important for personalized dose adjustments to improve the effect and safety of the therapeutics

and clinical utilization of genotypes of *CYP2D6* is increasing over the years (Molden et al., 2021).

1.2 Problem Statement

The Asian population is distinctly diverse from the Western populations. Although pharmacogenomic testing is available for several drugs, there is a limitation that it might not be applicable in all populations. A considerable portion of medicines commonly prescribed are metabolized by enzymes coded by highly polymorphic genes (Finkelstein et al., 2016). In Asia, North-east Indian populations showed a genetic affinity with Mongoloids from southeast Asia. North Indians and north-eastern populations are markedly unrelated. The Northeast Indian population exhibits significant genetic diversity compared to other regions (Agrawal et al., 2008). There could be the possible influence of genetic variations in drug responses in this population which is crucial for tailoring the treatment of the most prevalent diseases to individuals' needs. Despite understanding the importance of Pharmacogenes in invoking ADR, their polymorphisms are not studied in response to population level.

Furthermore, when examining cancer incidence data spanning from 2012 to 2016, sourced from the 11 Population-Based Cancer Registries (PBCRs), it becomes evident that the northeastern region of India carries the highest burden of cancer cases. Specifically, Aizawl and Kamrup Urban in Assam have consistently reported elevated cancer incidence rates since 2003, affecting both men and women alike. Of particular concern in this region is the prevalence of gastric cancer, which imposes a substantial healthcare challenge. Aizawl district in Mizoram stands out with the highest incidence of gastric cancer among men. Several risk factors contribute to this alarming trend, including Helicobacter pylori infection, advancing age, a diet high in salt, and the consumption of diets low in fruits and vegetables, as highlighted in studies by Shanker et al. (2021) and Chakraborty et al. (2021). Intriguingly, despite the pressing need for tailored treatment approaches in the context of gastric cancer in the Northeast Indian population, there is a noticeable gap in comprehensive studies on pharmacogenetics. This knowledge gap prompted the initiation of the present study, which aims to rectify this deficiency by providing valuable data on pharmacogenetics, thereby contributing to the advancement of more effective gastric cancer treatment strategies in this region.

1.3. Review of Literature

1.3.1 Reference genome and genome variation

In 2001, the Human Genome Project (HGP) revealed the human haploid nuclear genome as ~ 2.8 billion bp long consisting of nearly 30,000 to 40,000 protein-coding genes (Makałowski et al., 2001; Brown 2002). In 2004, the National Human Genome Research Institute (NHGRI), USA released another scientific statement by cutting down the number of genes to 20,000 to 25,000 genes while the majority of sequences are junk DNA and non-coding regions ("NHGRI History and Timeline of Events": https://www.genome.gov/about-nhgri/Brief-History-Timeline). Since then, the effort has been made to construct a complete reference genome of humans as the initial releases of the human genome were not error-free and certain regions were unable to sequence. Successive assembly of the human reference genome was released with various improvements by introducing alternate haplotypes (Guo et al., 2017).

According to GRCh (https://www.ncbi.nlm.nih.gov/grc/help/patches/), the current version of the reference genome assembly (GRCh38: Genome Reference Consortium Human build 38) was released in 2013 and since then various patches as a result of alternate haplotypes coming out of different genome projects were updated. After publishing the updated version of the reference genome, different consortiums kept on trying to improve the accuracy of the reference genome. Any new updates on the scaffold sequence are updated using patches. These patches are released from time to time. GRCh38 has been updated with patch 14 on 22 March 2022. These patches are necessary for updating the correct sequences and filling up the genome gaps. The exonic region also referred to as Exome is a major part of the genome as it is the coding region of the genome. Compared to GRCh37, GRCh38 is improved substantially in exonic sequence coverage and reducing pseudogenes. Another improvement in the GRCh38 was the inclusion of improved modelled centromeric sequences to fill the gap. However, due to repetitive sequences, it was still not a complete representation of the human genome and the major challenges in completing the telomeric region of the chromosomes. These repetitive sequences made it complex to map short reads and thus long read sequencing strategy was sought to avoid ambiguous mapping.

Previous reference genome assemblies were completely missing from the centromeric and pericentromeric sequences. In humans, the centromere is composed of Alpha Setelite DNA which is rich in Adenine and thymine. A complete gapless reference genome is extremely necessary for accurate variant calling. A gapless genome influences positively the discovery of SNPs, Structural variations, Fusion genes, and gene expression estimation. Development of Telomere-to-Telomere-CHM13 (T2T-CHM13) reference assembly have almost solved the long-standing issue of missing region in the genome using PacBio High Fidelity (Hi-Fi), Illumina Short-Read, Illumina Hi-C, BioNano Optical Map, Single Cell template strand sequencing and long read sequencing by oxford nanopore to assemble the nearly complete uniform homozygous genome from CHM13hTERT cell line (46, XX). However, this reference genome version is likely to represent European ancestry with Neanderthal admixture (Haas et al., 2013; Sergey et al., 2022). Due to the inclusion of a small number of individuals in the previous CHM genome, it is still not clear how accurately variations can be identified across populations. To address this problem, the human pan-genome reference consortium came up with the first draft of a complete pan-genome derived from 47 individuals across different populations. The pan-genome is likely to cover 350 individuals from genetically diverse backgrounds which is aimed to represent genetic variation across populations (Liao et al., 2023).

1.3.2 Single Nucleotide Polymorphisms

The reference genome acts as a baseline for comparing other genomic studies such as variant calling, gene expression study, and gapless genomes providing an opportunity to improve such analysis by mapping the genomics reads to newly discovered gene or chromosomal locations. The addition of 189 Mbp of sequences missing from previous assemblies is likely to improve variant calling using the T2T-CHM13 genome (Sergey et al., 2022; Nicolas et al., 2022). It has been studied that the Reference genome plays a crucial in genetic variation study in different populations. The evolution of the reference genome also triggered the discovery of new polymorphism; a genetic alteration that occurs in greater than 1% of the population that is responsible for variation in populations. One of the major applications of the reference genome is studying genetic variation across the

population. Nevertheless, GRCh37 (hg19) and GRCh38 have played important roles in the discovery of millions of Single Nucleotide Polymorphisms (SNPs), Structural Variations including gene fusions in humans globally (Brian et al., 2008).

It was well understood that the most common variations in human genetics were Single nucleotide polymorphisms. Simultaneous efforts were going on to study the genetic variation in different populations. In 1998, the dbSNP database was made public and centralized for SNP cataloguing, which currently hosts both SNPs and clinical mutations (Sherry et al., 2001). There was a paradigm shift towards the study of clinically relevant variations and mutations by different clinical diagnostic laboratories, and research groups and there was a need for evidence-based interpretation of the variants. In 2013, ClinVar was introduced with variants with clinical evidence by expert panels, researchers, and Clinical Laboratories (Landrum et al., 2008).

Over the years different genomics research projects were taken up to understand human genome variation using advanced DNA sequencing technologies. Moreover, another such major program was the 1000 Genome Project to catalogue genetic variation to represent global genetic variation in 26 populations using WGS and deep WES (1000 Genomes Project Consortium, 2015). Similarly, the UK's One Lakh genome Sequencing project aimed to sequence (WGS) in patients from the UK's National Health Service. The primary motive of genomic variant studies is to improve clinical care in different diseased patients (Samuel et al., 2017).

Various genome sequencing projects to understand cancer was inspired by the successful completion of the Human genome project. Human genome variation study revolutionized the way to understand infectious diseases such as Tuberculosis and SARS-COV2. Apart from the virulent factors present in the pathogens, there is a lot to know about the host genetics. Host genetics plays an important role in disease pathogenesis. There are population-specific variants in MHC Class II, *ESRRB, TGM6,* and *ASAP1* genes found to be associated with TB. Likewise, many gene variants (*HLA-DR*, *HLA-DQ*, and *TLR1*) have been identified that increase the susceptibility of *Mycobacterium leprae* in the Indian population. *Plasmodium falciparum* (causes

malaria) and hepatitis C virus (HCV) pathogenesis are also influenced by host genetics (Kwok et al., 2021).

Similarly, the discovery of genetic variants through next-generation sequencing has revolutionized cancer diagnosis, and treatment strategies as well as improved the understanding of cancer disease mechanisms. WGS and WES method implies sequencing of the whole DNA and the coding part of the DNA, respectively, and subsequently can detect mutation by comparing it to the reference genome. These genetic mutations occur in two ways, namely germline variants and somatic mutations. Germline mutation occurs in egg or sperm cells and passes through the generation making the progenies vulnerable to disease. These variants may occur as singleton or pass through as a group together, known as a haplotype. Moreover, the germline variants also share somatic mutations in cancer (Crawford et al., 2005; Meyerson et al., 2020). The type of variants causing cancer also depends on the age and onset of cancer in the patient. Most germline missense and pathogenic variants in oncogene and tumor suppressor genes may cause early onset of cancer depending on the zygosity. While somatic mutation needs extreme exposure to mutagens or external factors to occur and is likely to cause the late onset of cancer (Qing et al., 2020).

Identification of driver mutation in cancer has enabled the development of new diagnostic approaches and therapies (Gagan et al., 2015). Several databases catalogued cancer driver mutations, but catalogue of somatic mutations in cancer (COSMIC) and The Cancer Genome Atlas (TCGA) are the most widely used (Wang et al., 2016; Tate et al., 2019). A small number of mutations can cause cancer and develop resistance against therapies. These driver mutations have a wide range of clinical applications and could serve as biomarkers for cancer diagnostics and prognostics. Driver mutations may also contribute to the phenotype in the patients. However, the identification of passenger mutation (Mutations assumed to be neutral) is also important as these can be changed to driver mutation over time to exert a combinatorial effect. There are several driver mutations identified across cancer types in recent years and are associated with specific signatures, for example, G12C in *KRAS* mutation has been linked to smoking-related signatures in lung adenocarcinoma, R249S in *TP53* mutation has been linked to aflatoxin signature and V600E in *BRAF* has been found

linked to sun exposure in melanoma (Ostroverkhova et al., 2023). NGS can be extensively applied in identifying somatic mutations that could reveal the potential drivers (Martínez-Jiménez et al., 2020).

1.3.3 Pharmacogenetics and Pharmacogenomics

Studies on genetic variation are not only limited to the investigation of drivers or casual variants of a disease but can also be used in personalized medicine. Genetic variants also impact the prognosis of the disease if occurs in the genes involved in the drug absorption, distribution, metabolism, and excretion, commonly referred to as ADME (Laura et al., 2021). Such genes involved in ADME are also referred to as Pharmacogenes (PGx-Genes). These PGx-Genes are further categorized into drug-metabolizing enzymes, transporter, major histocompatibility complex (MHC) genes, and drug target genes (Johnson et al., 2001; Ingelman-Sundberg et al., 2005; Pavlos et al., 2011; Niemi et al., 2011). Pharmacogenomic strategies mitigate the trial-and-error paradigm in drug prescription, effectively limiting patients' exposure to medications that may prove ineffective or pose potential toxicity risks based on their individual genetic profiles (Katara et al., 2014). The common polymorphism in this gene may result in variable drug action. Variations in drug targets often develop resistance against a particular therapy. Similarly, variations in drug-metabolizing enzymes result in poor, intermediate, normal, and rapid metabolism of the drug. Pharmacogenomics primarily aims to identify risk alleles associated with efficacy and toxicity for improving better treatment regimens design and dose adjustments (Roden et al., 2019).

Pharmacogenetics and pharmacogenomics have played pivotal role in personalized medicine, prescribing drugs based on a patient's genetic profiles. The International Conference on Harmonisation, a global group of regulatory agencies, has described pharmacogenomics as the study of differences in DNA and RNA building blocks concerning how drugs work, and pharmacogenetics as the study of differences in DNA sequence concerning how drugs work. Pharmacogenetics explores how a person's genetic makeup influences how they respond to medicines, while pharmacogenomics examines the combined impact of multiple genetic mutations in the genome that can affect how a patient responds to drug treatment (Roden et al.,

2019). Dere et al. (2009) categorized ways in which genetic variants may alter responses to drugs by "1) variation in the metabolism of a drug among individuals; 2) variation among population members concerning drug adverse effects that are not based on the drug's action; and 3) response or lack response by genetic variation in the drug treatment target". Lately, many cases involving differences in how people respond to drugs have been reported. This has drawn the attention of the scientific community to the research on genes related to drug response, combining both experimental and bioinformatics methods. Currently, numerous research are being conducted to understand the variation of Pharmacokinetics (PK) and Pharmacodynamics (PD) related genes and their association with VDR (Krebs et al., 2019; Katara et al., 2019).

A genetic variation can be seen as a useful marker for guiding medical treatment if it changes how a gene works, can be affected by drugs that are already approved or being studied or can help predict how someone will respond to a particular drug or therapy, including their sensitivity, resistance, and risk of side effects. Toxicity to a specific drug/therapy may often hamper the prognosis of the disease either by improperly metabolizing drugs or by some adverse reactions (Bush et al., 2020). Pharmacogenomics aims to study the genetic variants that are responsible for drug response.

The PREDICT program of Vanderbilt University Medical Center has made significant efforts in using pharmacogenomics into day-to-day clinical practice. The ultimate aim of the program is to investigate the impact of common genetic variations in clinical genes affect drug responses. Pharmacogenomics involves identifying relevant gene variations in patients before they require specific medications, enabling healthcare providers to make better-informed decisions about drug choices and dosages (Van Driest et al., 2023).

1.3.4 Adverse Drug Reactions (ADRs)

Adverse Drug Reaction (ADR) was considered to be a global medical concern. In 2000, Edwards and colleagues defined ADR as a notably harmful or unpleasant response that might be arised from any medication. ADR indicates a potential risk if

the medication is administered again and should be either prevented, treated, or addressed by changing the dosage or discontinuing the product. Similarly, it was defined by the World Health Organization (WHO) as “a response to a drug that is noxious, unintended, and which occurs at doses normally used in man for prophylaxis, diagnosis or therapy of disease or the modification of a physiological function” (Sharma et al., 2014). In clinical setups, ADRs are considered major healthcare problems. However, patients with mild reactions are likely able to complete a given treatment course in particular cases as mild ADRs are easily manageable (Hacker 2009).

According to WHO, ADRs are classified into “six types (with mnemonics): dose-related (Augmented), non-dose-related (Bizarre), dose-related and time-related (Chronic), time-related (Delayed), withdrawal (End of use), and failure of therapy (Failure)” (Edwards et al., 2000). ADR monitoring is an important part of pharmacovigilance**,** however, the majority of ADRs are not reported and remain a major issue in health care (Coleman et al., 2016). Some ADRs can be predictable by studying drug pharmacokinetics and pharmacodynamics properties. Differences in specific PGx-genes can significantly impact an individual's drug response, which is a key factor in determining how effective the treatment is and the likelihood of experiencing adverse drug events. Some cumulative efforts have been put forward by many countries to study such gene variations and their effects on PK (Bush et al., 2016; Krebs et al., 2019).

Before FDA approval, drugs undergo rigorous clinical trials; however, these trials often lack the necessary scale to detect both rare and common serious adverse drug reactions (sADRs). Additionally, they frequently involve healthier subjects and adhere to an accelerated timeline, which limits their capacity to comprehensively identify sADRs. The post-approval surveillance systems, such as FAERS and MAUDE, suffer from irregular and incomplete reporting of sADRs that occur outside the clinical trial environment. This leads to delays in identifying and disseminating safety information. In response to these challenges, the Southern Network on Adverse Reactions (SONAR) was founded in 2010 as a pharmacovigilance initiative. SONAR's primary objective is to expedite the identification and sharing of sADR information,

reducing the current 11-year timeframe to 1-2 years. SONAR employs a methodical approach to pharmacovigilance, encompassing comprehensive medical reviews, case studies, extensive database analysis, policy assessments, and fundamental scientific investigations. The engagement of diverse experts and the inclusion of patient perspectives are pivotal in ensuring a holistic grasp of drug safety issues (Lu et al., 2014; Bennett et al., 2020).

In 2015, a total of 82 pharmacogenetic variations in ~9000 samples were reported by electronic Medical Records and Genomics (eMERGE) in collaboration with Pharmacogenomics Research Network (PGRN) using targeted sequencing (Bush et al., 2016; Krebs et al., 2019). Pharmacogenomics Knowledgebase (PharmGKB: https://www.pharmgkb.org/) has so far identified 149 pathways, 68 very important pharmacogenetics (VIPs) as well as 23,975 variant annotations. Moreover, it provides 4,568 clinical annotations (Level 1A and Level 1B) and 753 Drug Label Annotations. For example, PharmGKB ID PA166183593 reports about the *CYP2C9*2* or **3* variants that have reduced clearance of acenocoumarol (Sintrom). Apart from the previously mentioned resources, PharmVar (https://www.pharmvar.org/) and Pharmacogenetic Databases (http://www.pacdb.org/) are two additional critical databases in the field of pharmacogenomics. PharmVar serves as a comprehensive repository for genetic variants that influence how individuals respond to medications, offering researchers and healthcare providers a valuable reference for understanding the genetic factors behind drug reactions. Similarly, the Pharmacogenetic Databases provide a wealth of information about how specific genes can impact drug responses, acting as a vital resource akin to an encyclopaedia for pharmacogenomic data. Together, these databases play an essential role in advancing personalized medicine by equipping professionals with the knowledge needed to tailor drug treatments to individual genetic profiles, ultimately enhancing patient care and medication outcomes.

1.3.5 ADRs in Cancer and the Role of SNPs

ADRs are common and unavoidable risks associated with cancer chemotherapy. Detecting, monitoring, and preventing ADRs are crucial aspects of

cancer patient care. A prospective study in Nepal revealed that age over 60 and female gender were risk factors for ADR development due to anticancer medications. The primary ADR-causing drugs were alkylating agents and antimetabolites, with specific drugs like Carboplatin, Gemcitabine, and fluorouracil were playing significant roles (Shrestha et al., 2017). Anaemia was the most prevalent ADR, and many ADRs persisted even after discontinuing the suspected drug. Most ADRs were considered probable in causality, moderate in severity, and probably preventable. ADRs increase the cost of illness due to additional therapy, clinical investigations, and prolonged hospital stays. Managing ADRs remains a significant challenge in cancer patient care, necessitating vigilance, monitoring, and prevention to enhance pharmaceutical care for patients (Shrestha et al., 2017). Assi et al. (2020) analysed 110 threads from online discussion forums, identifying 473 ADRs primarily linked to the nervous and immune systems. The examination unveiled three primary themes: advice between patients, self-administered treatment, and alterations in lifestyle. It emphasized the importance of considering patient experiences and attitudes when designing treatment plans, highlighting the need for improved communication between healthcare professionals and patients.

The drug 5-Fluorouracil (5FU), which is used in cancer treatment, has been known to cause various side effects. Nausea and vomiting are among the most concerning side effects for cancer patients. The likelihood and intensity of these side effects can vary depending on factors such as the dose, treatment schedule, combinations with other medications, and individual patient characteristics. 5FU can also lead to ulcer-like sores along the digestive tract, which may manifest as mucositis, pharyngitis, esophagitis, gastritis, colitis, or proctitis. When 5FU is administered combined with leucovorin through continuous intravenous infusion, the chance of experiencing oral mucositis and diarrhoea is higher. Severe myelosuppression, which is a significant cause of illness and even death in cancer patients, is relatively rare when 5FU is used at the typical dose (Kown, 2005). Fortunately, there are various medications like selective 5-HT3 antagonists and corticosteroids available to help manage 5FU-induced nausea and vomiting. Additionally, colony-stimulating factors can be used to boost the production of white blood cells in the bone marrow. In general,

progress in supportive care has diminished the incidence of severe side effects caused by 5FU, enabling the preservation of scheduled dose intensity. This review addressed the adverse effects associated with 5FU-based chemotherapy, encompassing issues like nausea and vomiting, myelosuppression, as well as neuro-cardio-cutaneous toxicities (Kown, 2005).

Gianni et al. (2009) discussed the occurrence of Apical Ballooning Syndrome (ABS), a stress-induced cardiac condition characterized by transient abnormalities in the wall motion of the heart, in a woman who had recently undergone chemotherapy for metastatic colorectal cancer. She presented with chest pain, elevated cardiac enzymes, and ST-segment abnormalities on her electrocardiogram, but coronary angiography revealed no blockages in her coronary arteries. The heart condition of the patient was likely influenced by both the emotional stress of a cancer diagnosis and the chemotherapy treatment. However, the patient was fully recovered which was confirmed through echocardiography (ECG). This case pointed out the significance of recognizing and properly managing potential heart-related adverse events of chemotherapy. A research study compared the safety profiles of oral fluoropyrimidines with 5-Fluorouracil (5FU) by examining ADR reports submitted to the US-FDA Adverse Event Reporting System (AERS). After thoroughly reviewing 1,644,220 reports from 2004 to 2009, it was found that 5FU was more frequently associated with conditions like leukopenia, neutropenia, and thrombocytopenia, while capecitabine was linked to certain conditions such as diarrhoea, nausea, vomiting, and hand-foot syndrome. This particular study also suggested that the FDA's AERS and data analysis methods were valuable for identifying potential issues (Kadoyama et al., 2012).

In another study, researchers studied the occurrence of heart-related problems caused by 5-Fluorouracil (5FU) and capecitabine in patients with colorectal cancer and aimed to pinpoint associated risk factors. Among the 995 patients who received 5FU and 1241 who were given capecitabine, it was found that 5.2% of those treated with 5FU experienced heart-related adverse reactions, compared to 4.1% in the capecitabine group. Adverse events included angina without ischemia, angina with ischemia as confirmed on ECG, unspecified chest pain, ST-elevation myocardial infarction, and non-ST-elevation myocardial infarction were found most common

heart-related issues. There were a small number of cases involving cardiac arrest or sudden death. While the study did not identify specific risk factors for heart-related ADRs with 5FU, it did find that ischemic heart disease was a risk factor for heart problems induced by capecitabine. The research indicated that both 5FU and capecitabine could lead to heart-related issues in a small portion of patients and ischemic heart disease as a risk factor for capecitabine-induced heart problems (Dyhl-Polk et al., 2020).

In a recent case study, an unusual and severe reaction to Capecitabine was observed in a 70-year-old male patient with metastatic pancreatic cancer. The patient developed Steven-Johnson syndrome (SJS) after just ten days of Capecitabine treatment, experiencing symptoms such as vomiting, mucositis, hyperpigmentation, itching, and scrotal mucosal peeling (Karthikeyan et al., 2022). Despite medical intervention, the condition worsened, and the patient unfortunately passed away. This case highlights the importance of healthcare providers being aware of rare adverse effects, like SJS, associated with Capecitabine and other drugs. It emphasizes the need for educating and counselling patients about potential adverse effects to ensure early detection and intervention, ultimately reducing the risk of severe complications (Karthikeyan et al., 2022). Predicting the risk of severe adverse reactions to chemotherapy is vital for tailoring effective cancer treatments. While current guidelines recommend genotyping specific *DPYD* variants to assess toxicity risk associated with fluoropyrimidines, this approach falls short in clinical practice. In this study, novel genetic variants were sought by examining a set of tag SNPs in genes connected to fluoropyrimidine pharmacodynamics. The study included the genetic testing of 23 specific SNPs in a group of 301 patients with colorectal cancer who were undergoing capecitabine-based chemotherapy. Through both individual and combined statistical analyses, the researchers identified ten SNPs that were linked to severe adverse reactions to capecitabine. Notably, these included variants of *CDA*, *DPYD*, *TYMS*, *SLC22A7*, and *UMPS*. Notably, before this study, no link had been established between these SNPs and capecitabine-induced toxicity, demonstrating the effectiveness of tag SNPs in uncovering previously unknown polymorphisms. These findings suggested that variants hold the potential to enhance the predictive accuracy

of existing tests, thereby reducing the occurrence of severe adverse reactions to capecitabine (Pellicer et al., 2017).

A study was conducted by Etienne-Grimaldi et al. (2017) to assess the relationship between *DPYD* gene variants and toxicity in advanced breast cancer patients receiving capecitabine, a fluoropyrimidine chemotherapy. Out of the 243 patients analyzed, 10.3% experienced grade 3 and 2.1% of grade 4 capecitabine-related digestive, neurologic, and hepatotoxicity. Exhaustive *DPYD* exome sequencing revealed 48 single-nucleotide polymorphisms (SNPs), including 19 in coding regions, with three novel variations, F100L and A26T being pathogenic. The presence of a specific set of harmful variants (*2A, I560S, and D949V) demonstrated a notable correlation with grade 3-4 toxicity, and the incorporation of supplementary variants (D342G, S492L, R592W, and F100L) enhanced sensitivity for both grade 3-4 and grade 4 toxicities. The study suggests that an extended set of deleterious *DPYD* variants enhances the predictive performance of *DPYD* genotyping, surpassing conventional genotyping limited to consensual variants. However, combining genotyping and phenotyping did not substantially improve sensitivity but compromised positive predictive value and relative risk.

Numerous paediatric cancer patients grappled with notable chemotherapy-related adverse effects. The likelihood of experiencing these drug reactions is influenced by variations in SNPs. The tangible impact of pharmacogenetics research on clinical outcomes in paediatric oncology is exemplified through the identification of genetic variations within the thiopurine methyltransferase (*TPMT*) gene. The genetic diversity observed in TPMT enzyme activity is predominantly represented by three distinct variants: TPMT2 (G238C), TPMT3A (G460A and A719G), and TPMT*3C (A719G), collectively accounting for over 95% of the variations (Conyers et al., 2018). This approach has become widely accepted as the standard procedure in medical practice for identifying genetic variations in *TPMT* among paediatric cancer patients at the initiation of therapy. This genetic screening enables healthcare providers to proactively adjust medication dosages. *TPMT* has emerged as the primary pharmacogenetic marker for tailoring drug dosages in cases of acute lymphoblastic

leukaemia (ALL), following established guidelines based on an individual's *TPMT* genotype (Conyers et al., 2018).

1.3.6 Metabolism of Chemotherapeutic Drugs

5-Fluorouracil (5FU) is commonly used chemotherapeutic administered intravenously with over 80% of it undergoing metabolism in the liver (Diasio et al., 1989). Capecitabine, an oral prodrug of 5FU, traverses the gastrointestinal wall unchanged and is subsequently transformed into 5'-deoxy-5-fluorocytidine and then 5'-deoxy-5-fluorouridine within the liver through the actions of carboxylesterase and cytidine deaminase, respectively (Miwa et al., 1998; Yen-Revollo et al., 2008). 5'dFUR is converted into 5FU with the assistance of thymidine phosphorylase or uridine phosphorylase (Miwa et al., 1998). Another prodrug of 5FU, tegafur, undergoes conversion by *CYP2A6* into an unstable intermediate known as 5-hydroxytegafur, which subsequently breaks down spontaneously to produce 5FU (Yen-Revollo et al., 2008).

The conversion of 5FU into dihydrofluorouracil (DHFU) occurs, and DHFU is further metabolized into fluoro-beta-ureidopropionate (FUPA), eventually leading to the formation of fluoro-beta-alanine (FBAL) (van Kuilenburg et al., 2003). These metabolites are eventually excreted from the body. Crucial genes involved in this pharmacokinetic pathway include dihydropyrimidine dehydrogenase (*DPYD*), dihydropyrimidinease (*DPYS*), and beta-ureidopropionase (*UPB1*), all of which play vital roles in the catabolism of 5FU.

1.3.7 5-Fluorouracil , Leucovorin and Oxaliplatin Toxicity Associated Variants

In a study of advanced gastric cancer patients receiving 5-Fluorouracil, leucovorin, oxaliplatin, and docetaxel (FLOT) chemotherapy, researchers identified a significant association between thymidylate synthase (TS) group A (low expression group) genotypes with two to three repeats (R) “(2R/2R, 2R/3RC, 3RC/3RC)” and an increased risk of grade 3/4 hematotoxicity. Specifically, 59% of TS-group A patients experienced severe hematotoxicity compared to 25% of TS-group B (High expression group) patients. TS-promoter polymorphisms were found to be potential markers for hematotoxicity in this patient group (Goekkurt et al., 2009).

Similarly, a recent study has found that cancer treatment with the chemotherapy combination 5-Fluorouracil , leucovorin, oxaliplatin, and bevacizumab (FOLFOX) resulted in a rare occurrence of coronary vasospasm, a form of cardiac toxicity. This adverse effect is unusual, especially in a patient who had previously tolerated 5-Fluorouracil alone without complications. The case highlights the potential for unidentified cardiovascular toxicities associated with chemotherapy regimens, particularly those involving FOLFOX and bevacizumab. The patient's successful treatment with dihydropyridine calcium channel blocker therapy suggests a potential avenue for managing such vasospasm induced by chemotherapeutic agents, adding to the understanding of chemotherapy-related toxicity (Kabir et al., 2020).

A comprehensive review citing variants in genes associated with toxicity against combination therapy of Fluoropyrimidine drugs, leucovorin and Oxaliplatin in different types of cancer has been provided in Table 1.1.

Table 1.1: Genetic variants significantly associated (P-Value< 0.05) with toxicity against combination therapy of Fluoropyrimidine drugs, leucovorin and Oxaliplatin in different types of cancer.

dbSNP_ID	Gene	Description	Citation
rs13181	*ERCC2*	The "G" allele is linked to a higher risk of experiencing drug toxicity when individuals with colorectal neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin, as opposed to those with the "TT" genotype.	PMID:20385995 (Voige et al., 2010)
rs17376848	*DPYD*	The genotype "AG" is associated with a greater severity of drug toxicity when individuals with colorectal neoplasms are exposed to capecitabine, irinotecan, and oxaliplatin, compared to those with the "AA" genotype.	PMCID: PMC4574842 (Falvella et al., 2015)

rs1801133	*MTHFR*	The "AG" genotype is linked to a higher risk of experiencing drug toxicity when individuals with colorectal neoplasms are treated with capecitabine, fluorouracil, leucovorin, and oxaliplatin, as compared to those with a different genotype.	PMID:20819423 (Kristensen et al., 2010)
rs45445694	*TYMS*	The genotype (CCGCGCCACTTCGCCTGCCTCCGTCCCG)2/(CCGCGCCACTTCGCCTGCCTCCGTCCCG)2 is associated with an increased risk of drug toxicity when individuals with colorectal neoplasms are treated with capecitabine, fluorouracil, leucovorin, and oxaliplatin, as compared to other genetic variations..	PMID:20819423 (Kristensen et al., 2010)
rs1801131	*MTHFR*	The genotype "GG" is linked to a higher risk of drug toxicity when individuals with colorectal neoplasms are treated with capecitabine, fluorouracil, leucovorin, and oxaliplatin, compared to other genotypes.	PMID:20819423 (Kristensen et al., 2010)
rs67376798	*DPYD*	Genotype AT is associated with increased risk of hand-foot syndrome when treated with bevacizumab, capecitabine, cisplatin, docetaxel, epirubicin, oxaliplatin or trastuzumab in people with Stomach Neoplasms as compared to genotype TT.	PMID:27995989 (Meulendijks et al., 2017)
rs25487	*XRCC1*	The "CC" genotype is associated with a lower risk of experiencing Peripheral Nervous System Diseases when individuals with colonic neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin, in	PMID:23314736 (Lee et al., 2013)

		contrast to individuals with the "CT" or "TT" genotypes.	
rs1695	*GSTP1*	The "GG" genotype is linked to a greater severity of Neurotoxicity Syndromes when individuals with colorectal neoplasms are treated with oxaliplatin, compared to those with the "AA" or "AG" genotypes.	PMID:17401013 (Ruzzo et al., 2007)
rs1695	*GSTP1*	The "AA" genotype is linked to a higher risk of experiencing Neurotoxicity Syndromes and Neutropenia when individuals with colorectal neoplasms are treated with fluorouracil, irinotecan, or oxaliplatin, in contrast to those with the "AG" or "GG" genotypes.	PMCID:PMC2903324 (McLeod et al., 2010)
rs1801133	*MTHFR*	The "AA" genotype is linked to a higher risk of Neutropenia when individuals with colonic neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin, as compared to those with the "AG" or "GG" genotypes.	PMID:23314736 (Lee et al., 2013)
rs1799794	*XRCC3*	The "CC" genotype is linked to a greater severity of drug toxicity when individuals with colorectal neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin, compared to those with the "CT" or "TT" genotypes.	PMCID:PMC3859145 (Cecchin et al., 2013)
rs717620	*ABCC2*	The "CC" genotype is linked to a higher risk of drug toxicity when individuals with colonic neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin,	PMID:23314736 (Lee et al., 2013)

		compared to those with the "CT" or "TT" genotypes.	
rs11615	*ERCC1*	The "AA" genotype is linked to a higher risk of developing Neutropenia when individuals with colonic neoplasms are treated with fluorouracil, leucovorin, and oxaliplatin, as compared to those with the "AG" or "GG" genotypes.	PMID:23314736 (Lee et al., 2013)

1.4 Chapterization of the study

Extensive review on the adverse drug reactions cancer therapy has found the lack of comprehensive studies on pharmacogenetics in the Northeast Indian population, especially concerning gastric cancer treatment despite of high stomach cancer incidence and mortality. This study aimed to bridge that gap, contributing valuable data to the field.

The current research is structured into the following six chapters:

- Chapter 1: Introduction and Review of Literature
- Chapter 2: Aims and Objective
- Chapter 3: Materials and Methods
- Chapter 4: Results
- Chapter 5: Discussion and Conclusion
- Chapter 6: Incidental Finding
- Chapter 7: Summary

CHAPTER 2
AIMS AND OBJECTIVES

2.1 Aims and Objectives

Despite many research efforts, it has been identified as a challenging task to monitor drug alteration and Adverse Drug Reactions due to genetic variation in cancer patients. This study aimed to understand the genetic variation of the genes responsible for all forms of drug metabolism and also the genetic variants responsible for ADR associated with Gastric Cancer.

The objectives of the proposed study are:

1. Cataloguing the coding and non-coding Pharmacogene variants in the north-east Indian healthy population.
2. Identification of the clinically actionable Pharmacogene variants in the Mizo healthy population.
3. Adverse Drug Reactions (ADRs) in Gastric Cancer (GC) patients and identification of genetic variants.

CHAPTER 3
MATERIALS AND METHODS

3.1 Recruitment of healthy volunteers

In the current research, a group of 93 healthy volunteers was carefully selected for a whole genome sequencing (WGS). The sampling process was started after the approval of the "Institutional Ethics Committee (IEC), Civil Hospital Aizawl, Mizoram (No.B.12018/1/13-CH(A)/IEC/94)". This study was a small part of the larger "Genomics for Public Health in India (IndiGenomes)" project which was initiated at CSIR-IGIB, New Delhi. To maintain the study's focus on the indigenous populations of the Northeast, individuals who resided in the region but did not have ancestral roots, were excluded from the sampling. The age range of the recruited volunteers were from 22 to 60 years.

Similarly, 27 self-declared healthy volunteers from various regions of Mizoram, belonging to the Mizo tribe were recruited for whole exome analysis. These individuals were genetically unrelated, ensuring diversity in the study sample. Ranging in age from 22 to 60 years, the volunteers were randomly selected from within the Mizo ethnic group. Prior to their participation, these volunteers provided informed consent, indicating their willingness to be part of the study after being made aware of its objectives and potential risks. The sampling process was started after the approval of the "IEC, Civil Hospital Aizawl, (No.B.12018/1/13-CH(A)/IEC92)". These approaches aimed to investigate genetic variations or mutations in pharmacogenes in the region and add useful information on genetic factors relevant to the health and well-being of the Mizo tribe in Mizoram.

3.2 Recruitment of gastric cancer (GC) patients

In this study, the patients were identified after cancer diagnosis by oncologists and pathologists and were approached with consent form and structured questionnaires after the approval of the IEC of Civil Hospital Aizawl (No. B.12018/1/13-CH(A)/IEC). A total of 59 gastric cancer patients diagnosed between 2016-2022 were identified from various hospitals in Aizawl, Mizoram including Civil Hospital Aizawl, Ebenezer Hospital, Aizawl Hospital, and Green-Wood Hospital, Mizoram. To conduct genetic

analysis, approximately 3 ml of peripheral blood was taken from each patient and preserved in EDTA vials and stored in deep freezer. The patients selected for inclusion in this study exhibited the age range from 31 to 85 years, which allowed for a broader exploration of the disease across different age groups. Additionally, data like demographic details and clinical data such as histological grading type, TNM staging, and treatment regimens were collected usinge structured questionnaire..

3.3 Adverse Event Data Collection

The research was aimed to investigate genetic variations and to understand their probable affect on drug metabolism which may exert ADR in the patients undergoing chemotherapy for stomach cancer at the state. A two-fold study was conducted to achieve this: First, the study initially focused on assessing the genetic variation among the patients receiving chemotherapy for stomach cancer. It aimed to identify genetic factors that might influence the response to widely prescribed chemotherapy. Secondly, a follow-up study was conducted specifically to monitor and document ADRs experienced by gastric cancer patients who had undergone chemotherapy and grading was performed using the guidelines of Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Out of the total number of patients who had received chemotherapy, 37 were successfully traced and followed up until 03 May 2023. This follow-up study was essential to comprehensively evaluate the safety and tolerability of the chemotherapy regimens used in clinical practice.

The ethical aspect of the research was upheld through approval by the IEC, Mizoram State Cancer Institute, Aizawl (D.12016/2/2013-MSCI/IEC/). Patients who were eligible for participation in the follow-up study were provided with a structured questionnaire designed to systematically document any potential adverse reactions resulting from their chemotherapy treatment. These patients were also required to provide informed consent, ensuring their willingness to participate in the study and share their medical experiences.

All adverse reactions in the GC patients were categorized according to established standards outlined in CTCAE, a part of the Cancer Therapy Evaluation Program (CTEP) of the National Cancer Institute, USA. This standardized

categorization ensured consistency in reporting and analyzing adverse reactions across the patient population. As part of the longitudinal aspect of the study, patients were followed up every six months to record their survival status. This monitoring task allowed for the calculation of their survival period, providing valuable insights into the long-term outcomes of gastric cancer patients who had received chemotherapy at MSCI.

3.4 Genomic DNA Isolation

The samples collected for the whole genome were processed for genomic DNA isolation at CSIR-IGIB, New Delhi where genomic DNA was extracted using the salting out method and subsequent quality control measures were followed before sequencing. Whole-genome sequencing (WGS) libraries were generated utilizing the following the manufacturer's protocol of "TruSeq® DNA PCR-Free Sample Preparation Kit" (Illumina Inc., San Diego, CA, USA, Cat. no. FC-121-9006DOC). Following the library preparation, paired-end sequencing of 150 bp read length was performed on the Illumina NovaSeq 6000 platform (Illumina Inc. San Diego, CA, USA) (Jain et al., 2021).

Similarly, the samples for WES were processed in an in-house laboratory for genomic DNA isolation. Genomic DNA had been taken from blood samples using the QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN). The manufacturer's instructions were followed in this process. After isolation, the genomic DNA was stored at -20°C for later use. In a microcentrifuge tube (1.5 ml), approximately 20 µl of QIAGEN protease enzyme was taken followed by addition of about 200 µl of the blood sample and equal amount of Buffer AL were added, and mixed by vortexing. The tube was kept for incubation for 10 minutes at 56°C and inverted five times. Any drops on the lid were removed using centrifugation followed by the addition of 200 µl of chilled ethanol and vortexing. After removing the drops from the lid by centrifugation, the entire lysate was moved to a mini-Spin column and placed in a collection tube of size 2 ml which were provided with the kit. This transfer was done using a pipette. The lid of the spin column was gently closed, and it was centrifuged at 8000 rpm for 1 minute at room temperature (RT). The flow-through was disposed off, and the spin column

was repositioned inside a fresh collection tube of size 2 ml. Approximately, Buffer AW1 of 500 µl was introduced into the mini-spin column and centrifugation was performed at ambient room temperature for 1 minute at 8000 rpm.

After discarding the flow-through, the spin column was kept in a new collection tube of 2 ml size. In the column, buffer AW2 of 500 µl was added and centrifuged at room temperature for 3 minutes at 14,000 rpm and the resulting liquid was disposed. The column was put back in the same collection tube for a 1-minute centrifugation just to ensure all liquid was removed. The 2 ml tube was replaced with a new 1.5 ml microcentrifuge tube and the column's caps were opened to air dry for 15 minutes in a Biosafety Cabinet. Approximately, 75 µl of AE buffer was added directly to the column's membrane, and it was left to incubate for 5 minutes at room temperature. Afterward, centrifugation was performed at 8000 rpm for 1 minute at room temperature to extract the DNA, which was stored at -20°C until ready for use in Whole Exome Sequencing.

About 0.64 g agarose powder and 80 ml of Tris acetate- EDTA (TAE) buffer was mixed in 100 ml of Conical flask and was boiled in the oven and cooled down. About 8 µl Ethidium Bromide (EtBr) was added to the Luke warm gel and poured on the tray and allowed to solidify for 40 minutes. The agarose gel after solidification was electrophoresised by loading 3 µl Genomic DNA and 2 µl 100 bp ladder to the well and run for 30 minutes to check the quality and concentration for Genomic DNA (Figure).

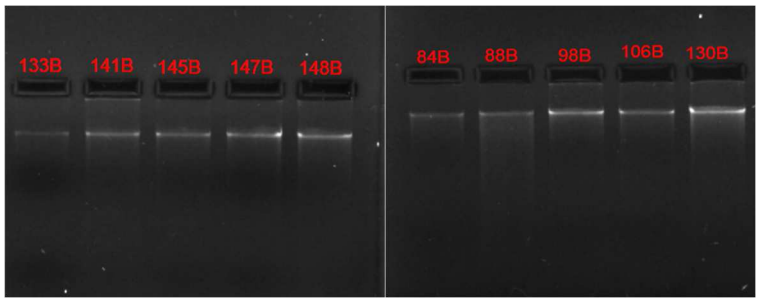


Figure 3.1: Agarose gel electrophoresis of DNA from representative samples

3.5 Whole Genome Sequencing and Data Analysis

Whole genome sequence data pre-processing, alignment to reference genome and variant calling was made using Illumina DRAGEN v3.4 (Jain et al., 2021). The joint variant calling for 93 individuals was performed using the Sentieon pipeline that integrates GATK protocols for joint genotyping which calculates the, allele count, allele number and allele frequency for the variants. (Jain et al., 2021).

3.6 Whole Exome Sequencing and Data Analysis

In this research project, the library preparation and whole exome sequencing (WES) processes were conducted in two distinct batches. The division was based on the timeline of sample collection. The samples collected during the period from 2016 to 2019 were processed and sequenced at the National Institute of Biomedical Genomics located in Kalyani, India. These samples were sequenced in the NIBMG, Kalyani in a collaborative effort. Agilent SureSelect^{XT} target enrichment System was utilized for WES library preparation using manufacturer protocol and sequening was performed in Illumina Hi-Seq2500 at 100X depth.

Samples collected after the year 2019 were processed and sequenced at Neuberg Diagnostics, Ahmedabad, India. This indicates a shift in the sequencing facility for the later samples. Neuberg Diagnostics provides more up-to-date sequencing technology and services. Twist protocol was used for WES library preparation using manufacturer protocol and Illumina NovaSeq6000 platform was used for sequencing at 100X depth. The overall workflow for sample Collection, data collection, WES, and data analysis has been explained in Figure 3.2 .

This division into two batches may have been influenced by various factors, including changes in technology, institutional collaborations, or the need for increased sequencing capacity as the research project progressed. Ultimately, it allowed for the efficient processing and analysis of samples collected over different time periods while ensuring that the most appropriate resources and expertise were applied to each set of samples.

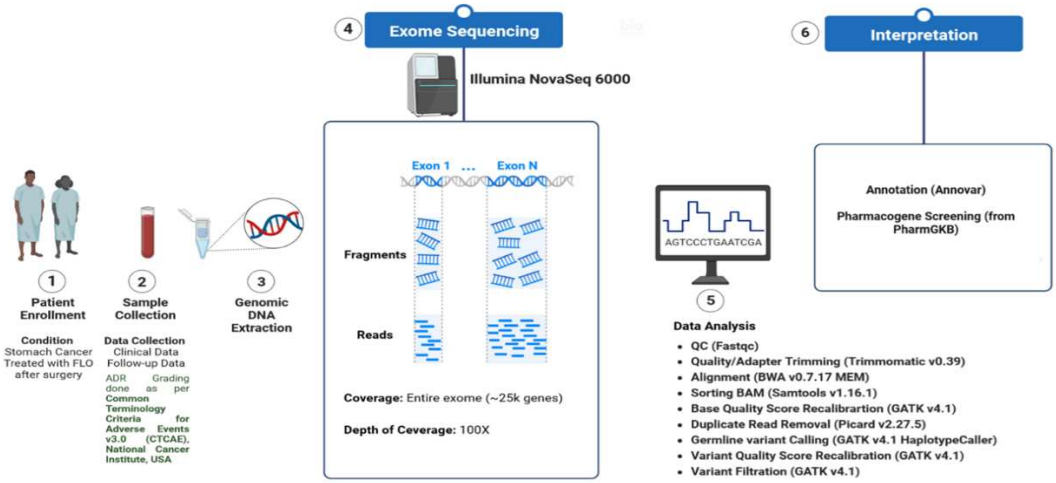


Figure 3.2: Overall workflow for Sample Collection, Data Collection, WES, and Data Analysis

3.7 Bioinformatics pipeline for Germline Variant Calling

WES data analysis was done using an HP Hi-End computing server, employing an internally developed automated pipeline known as WEAP (Whole Exome Analysis Pipeline). WEAP was specifically created for variant calling from trimmed FASTQ data and was designed to call both germline and somatic variants following the best practices outlined by GATK guidelines. It operates in two modes: serial mode, processing samples one at a time, and parallel mode, handling four samples simultaneously. Significantly, WEAP's parallel mode proved to be much faster when dealing with extensive sample sets. The pipeline WEAP incorporates various essential tools, including BWA aligner, Samtools, Picard, GATK, bedtools, vcftools, and Annovar, thereby establishing it as a comprehensive and standard automated solution for variant calling from WES data (Figure 3.3).

Data QC: To pre-process the raw WES reads obtained from the sequencing vendor, we developed an automated tool called QTQ (QC-Trimm-QC). QTQ is designed to handle a large number of samples efficiently by performing quality checks using Fastqc before trimming, trimming low-quality bases using the fastp tool, and then conducting a final QC assessment on the trimmed data (Andrew, 2010; Chen et al., 2018). QTQ is publicly available on GitHub at the following URL: https://github.com/ranjanjs34/QTQ. Certain samples were also trimmed using

Trimmomatic (Bogler et al., 2014) . The variant calling from WES data workflow involved a series of intricate steps to extract meaningful genetic information from whole exome sequencing (WES) data.

Data Alignment: Initially, the trimmed WES data, which had already undergone quality control checks were aligned to the GRCh38.p13 reference genome using BWA-MEM (Li, 2013). This alignment process generated SAM (Sequence Alignment/Map) files, which essentially provide information about how the sequencing reads map to the reference genome. To make the data more manageable and efficient, the SAM files were converted into the BAM (Binary Alignment/Map) format. Furthermore, the BAM files were sorted based on their genomic coordinates using the samtools utility (Li et al., 2009). This sorting facilitates subsequent analysis steps.

PCR Duplicate Handling: To ensure data accuracy, it's important to deal with PCR duplicates, which are multiple identical reads originating from the same DNA fragment. The Picard tool was employed to either remove these duplicates or mark them for later identification. The picard tool is available at github: https://github.com/broadinstitute/picard.

Base Quality Score Recalibration (BQSR): Accurate base quality scores are crucial for reliable variant calling. BQSR was performed using the GATK Baserecalibrator. It used a known variant dataset, "dbSNP138.vcf," to recalibrate base quality scores, correcting for systematic errors in the data.

Germline Variant Calling: The GATK HaplotypeCaller was employed to identify germline variants (genetic variations present in the patient's inherited genome) in gVCF (genomic Variant Call Format) mode. This mode allows for efficient storage of variant information (Van der Auwera et al., 2020).

Joint Genotyping: The data from both the GC samples (n=59) and the healthy control samples (n=27) were combined through joint genotyping using GATK's genotypegvcf tool. This step is essential for understanding how variants are distributed across the entire sample population (Van der Auwera et al., 2020).

Variant Quality Score Recalibration (VQSR): To refine the quality of the identified variants, VQSR was applied using the GATK VariantRecalibrator. This process helps to distinguish true variants from sequencing artifacts and noise (Van der Auwera et al., 2020).

Allele Frequency Calculation: Joint genotyping of the studied population was performed to calculate the allele frequency (AF) of variants. The prevalence of different genetic variations within the sample group can be best understood through AF . The allele frequency was counted by the formulae:

*Allele Frequency **(AF)**=Allele count **(AC)**/Allele Number **(AN)***
AN: Total number of individual x2

Hard Filtering: Variants that passed default hard filtering criteria were retained, while those failing these criteria were discarded. This hard filtering ensured that only high-confidence variants were used for downstream analysis. This hard filtering removes the low mapping quality variant, strand biasness based on different statistial test (Van der Auwera et al., 2020).

Variant Annotation: The annotated variants were enriched with additional information by the ANNOVAR tool. ANNOVAR uses various public databases such as RefGen, avsnp15, esp3500, ClinVar, ljb26, gnomad211_exome, exac03, and 1000g. Additionally, it provided functional predictions for the variants, integrating insights from tools such as SIFT, PolyPhen, MutationTatser, MutationAssessor, and others (Wang et al., 2010).

This entire process of aligning the data to annotating the variants was executed using the WEAP version 1 pipeline (Figure 3.3). Additionally, a user-friendly manual for germline and somatic variant calling with WEAP is publicly accessible on GitHub at https://github.com/ranjanjs34/WEAP. This comprehensive analysis pipeline ensures that high-quality genetic variant data is obtained from WES data, enabling researchers to investigate genetic variations in the studied population effectively.

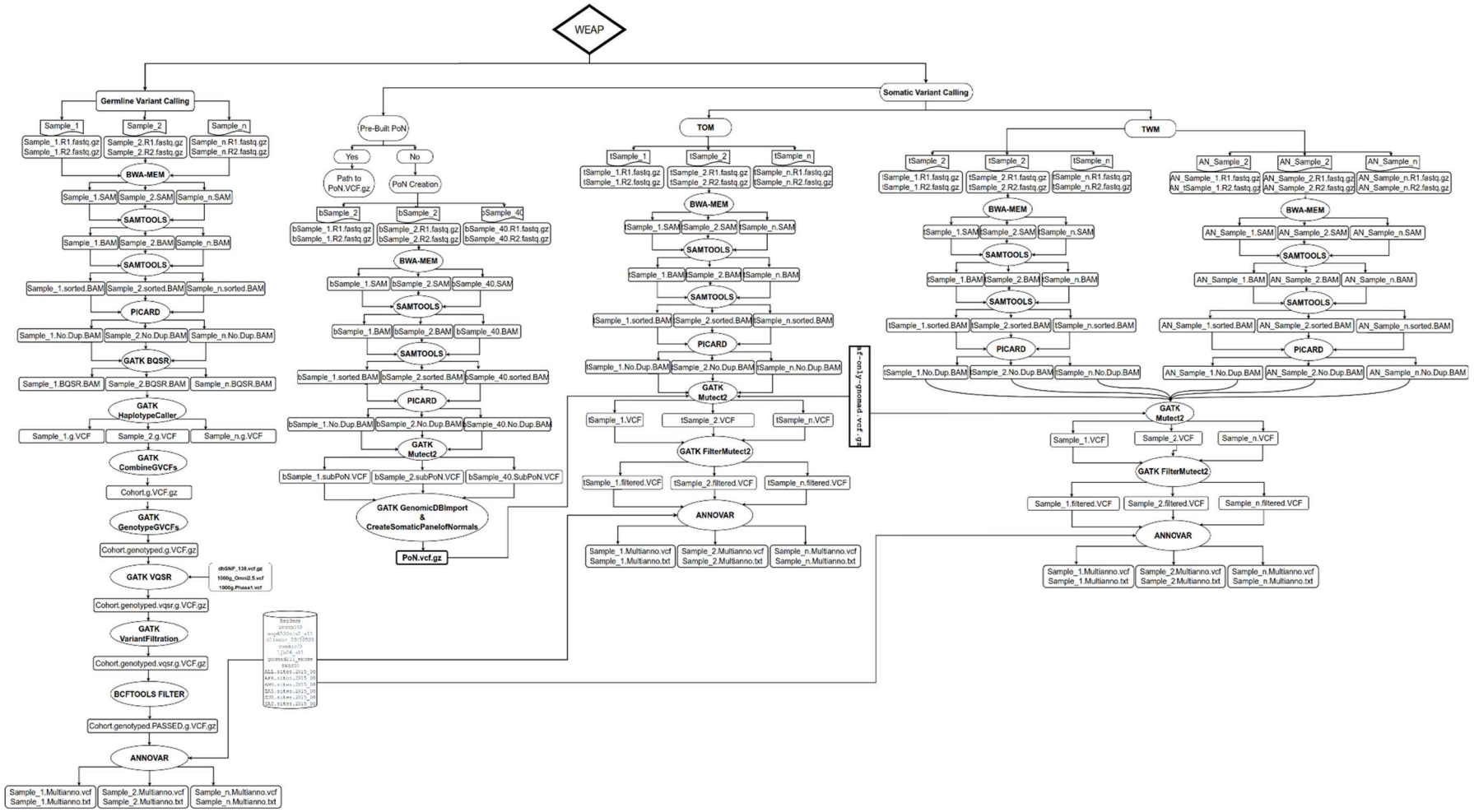


Figure 3.3: Overall workflow of WEAP Pipeline for Germline and Somatic Variant calling using GATK

3.8 Cataloguing genetic variants of VIPs

VIPs, which are genes of paramount importance in pharmacogenomics, were sourced from PharmGKB (https://www.pharmgkb.org/vips, accessed: 1 May 2022. PharmGKB serves as a comprehensive repository encompassing genes, their various genetic variants, star alleles, phenotypic information, clinical guideline annotations, drug label annotations, clinical annotations, and variant annotations. Additionally, it includes data on relevant pathways and supporting literature. These essential genes, designated as VIPs, were meticulously extracted from PharmGKB and organized into a text file for further analysis and research purposes.

In order to accurately retrieve genetic variants from the previously annotated results, we have developed a Python3-based tool called VariantExtractor, which is accessible at https://github.com/ranjanjs34/VariantExtractor . This tool accepts the gene list of interest as input from a text file and extracts genetic variants from all the data from the samples and writes the output into files in sorted (by gene name) order. Subsequently, all the extracted variants were combined and duplicate hits were removed to ensure only the unique variants. The distribution of variant types was calculated based on their functions, proportion of synonymous and nonsynonymous variants, as well as based on their genomic locations, including intronic, exonic, UTR region, and intergenic regions.

3.9 Development of Pharmacogenomic Variation Database

In order to provide a improved accessibility to gene variants within both whole exome and whole genome datasets, structured query language (SQL) was used. This strategy led to the development of a dedicated database known as MPVardb version 1. This database was carefully designed using MariaDB, Hypertecxt Preprocessor (PHP), HyperText Markup Language 5 (HTML5) with HyperText Markup Language (CSS), and was hosted on an Apache web server using XAMPP environment. The development of the database enhanced the efficiency and effectiveness of retrieving variants of the gene of interest from these datasets. The allele frequency calculated from the whole exome datasets were also incorporated to each variant. The final datasets were imported to the MariaDB database. The database was connected to a

webpage designed using HTML and PHP that accepts the query as a gene name in the hugo gene nomenclature system (HGNC). Visualization of the variant types from both datasets was done using Java script implemented in the PHP result page (Figure 3.4).

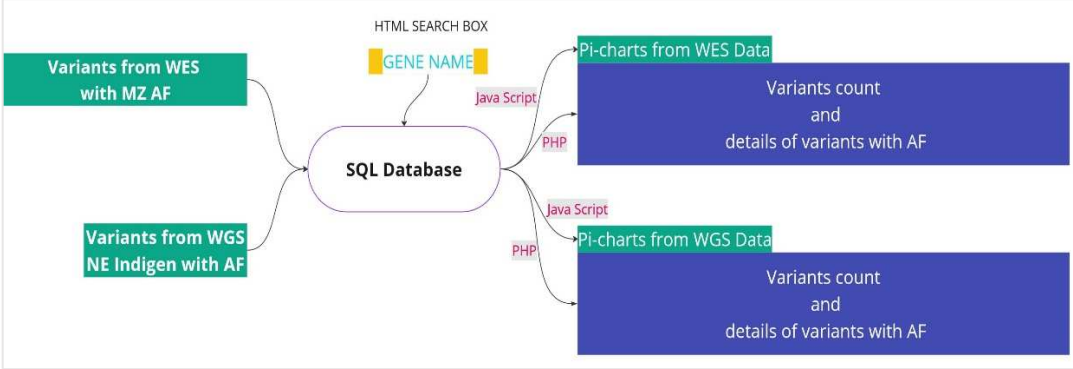


Figure 3.4: Workflow of construction of Mizoram Pharmacogenomic Variant Database bersion 1 (MPVardb V1)

3.10 Screening Actionable Pharmacogenes

The variants with drug dosing guidelines set by Pharmacogenetics Implementation Consortium (CPIC) and Dutch Pharmacogenetics Working Group (DPWG) were also screened (Sangkuhl et al., 2020). Moreover, the identification of important star alleles provides additional meaningful variant information. Understanding the specific variants comprising a haplotype, along with the diploid content in an individual, holds paramount importance in investigating drug metabolism, drug responsiveness, and adverse drug reactions (Hari et al., 2023).

To annotate the pharmacogenomic variant, the current research utilized the joint genotyped variants with score recalibrated and filtered by GATK derived from healthy exome dataset. The data was annotated using PharmCAT (https://pharmcat.org/; Sangkuhl et al., 2020). PharmCAT generated the clinical report based on the pharmacogene genotypes that matched with the prescribing recommendations which can be used to inform treatment decisions. To avoid the misinterpreting, the variants loci such as the chromosome with variant site , reference

allele and altered alleles wre crossed checked with the annotated data using ANNOVAR which was implementaed via. WEAP.

3.11 Adverse drug reaction (ADR) in Mizo GC patients

In our study, we employed a structured methodology to analyze adverse events in gastric cancer patients following chemotherapy. We first collected comprehensive data on the specific chemotherapy regimens administered to the patients, recording the number of individuals who received each regimen. Subsequently, we stratified the data by gender to ascertain the male-to-female ratio in relation to the occurrence of adverse drug reactions (ADRs). Additionally, we identified and documented the most prevalent ADR observed in our patient cohort. Furthermore, our analysis extended to examine the most common ADR with other concurrent ADRs, providing a deeper understanding of the between adverse reactions and other health conditions among these patients.

3.12 ADR and genetic variants analysis

In the methodology, it was recognized that a majority of adverse reactions stem from genetic variants affecting drug metabolism enzymes. These variants render the enzymes less efficient, resulting in poor metabolization of drugs. Therefore, the most frequent chemotherapy regimen administered to gastric cancer patients in Mizoram was necessary. This initial step allowed us to elucidate the complexities of drug metabolism and identify the pivotal genes involved in this pharmacokinetic/pharmacodynamic (PK/PD) pathway. The detailed curated metabolic pathway was carefully studied in PharmGKB and the genes were retreived from the database. Subsequently, we screened the annotated datasets for genetic variants within these key metabolic pathway genes. The patient IDs showing positive for suspected variants were carefully chosen to analyze their ADRs and survival period.

Additionally, we conducted an extensive screening for variants displaying a significant correlation with toxicity ($p\text{-value}<0.05$) from published literature via. the PharmGKB database against the same chemotherapy regimen, which was given to the stomach cancer patients in Mizoram. This approach allowed us to pinpoint crucial

genetic factors contributing to adverse reactions associated with a specific chemotherapy regimen.

3.13 Expression Quantitative Trait Loci (eQTL) analysis for PK/PD genes

The PK/PD pathway genes were subsequently also investigated for their impact on protein expression due to the genetic variant. eQTL allows to check how a genetic variant may impact the protein expression. The analysis was performed using the publicly available dataset hosted in the Genotype-Tissue Expression (GTEx) portal. GTEx-eQTL dashboard (https://gtexportal.org/home/eqtlDashboardPage) was used to run the analysis where the variant of the PK/PD of genes was fed in the form of GRCh38 coordinates. The gene expression associated with their variants can explained by net effect size (NES) and Beta distribution-adjusted empirical p-values calculated by FastQTL. All the significant (p-value < 0.01) cis-eQTL (impact on the particular gene) were considered in this study.

3.14 Expression-based overall survival analysis

A deeper exploration into the significant genes identified in the eQTL analysis to investigate their impact on the overall survival of patients with stomach cancer. This exploration was conducted using the KM-Plotter tool (https://kmplot.com), focusing on patients who had received specific chemotherapy as adjuvant treatment. Our analysis included a total of 152 patients and 60 months of follow-up threshold was choosen. To provide the necessary gene expression data for this study, we leveraged datasets from the Gene Expression Omnibus (GEO), specifically GSE14210, GSE15459, GSE22377, GSE29272, GSE51105, and GSE62254)

3.15 PK/PD pathway genes mutation frequency in public Database

To investigate the mutation frequencies of genes within the PK/PD pathway of chemotherapeutic drugs, we utilized publicly accessible datasets, including those from TCGA (The Cancer Genome Atlas), accessed through cBioPortal (https://cBioPortal.org). The analysis encompassed five distinct datasets, specifically Gastric cancer (OncoSG, 2018), as well as four datasets related to STAD (Stomach Adenocarcinoma), which include Pfizer and UHK (University of Hong Kong, Nat Genet 2014), TCGA (Firehose Legacy), U Tokyo (University of Tokyo, Nat Genet

2014), and UHK (University of Hong Kong, Nat Genet 2011). Our objective was to gain insights into the occurrence of mutations within these genes across these diverse datasets.

3.16 Mutation-based overall survival analysis

Finally, an assessment was done of overall survival, taking into account the mutation status of genes within the PK/PD pathway associated with chemotherapeutic drugs. This analysis was conducted using the TCGA STAD dataset, and the tool employed was https://tcga-survival.com/. This step allowed to gain valuable insights into the impact of gene mutations on survival outcomes in the context of STAD.

CHAPTER 4

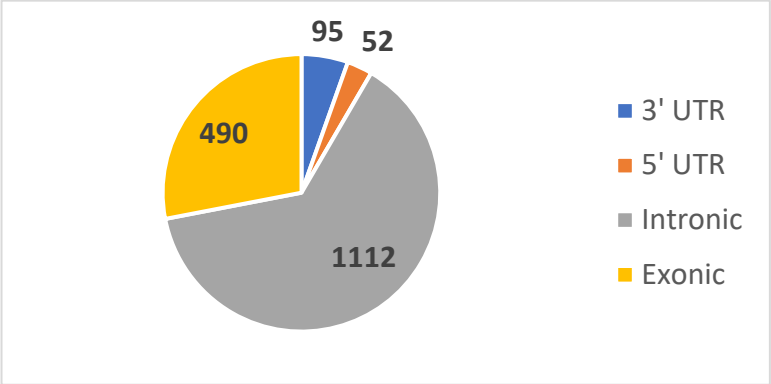
RESULTS

4.1 Catalouging PGx-Genes Variations in Healthy Individuals

A total of 1749 PGx variants of 67 VIPs passed the filtering criteria post-variant annotation. The allele frequencies were added to each variant calculated in the joint genotyping of all the 27 healthy exomes. These variants were from different regions of the exons covered by the sequencing reads. There were 490 variants within the coding, 1112 in intronic, 95 in 3'UTR, and 52 in 5'UTR regions. Among the variants, 163 variants were non-synonymous, and 315 variants were found to be synonymous. Moreover, 1 frame-shift substitution, 4 start-loss, 1 stop-loss, and 2 stop-gain (2) variants were found (Figure 4.1 A).

Similarly, about 77,799 variants were found in 67 VIPs in the NE IndiGenomes datasets. The AF was based on 93 healthy volunteers from different regions of northeast India. There were 2527 variants in the exonic region whereas 67,144 variants were found in intronic regions. There were 2086 variants in 3'UTR3, 378 variants in 5'UTR, 107 variants in intergenic, 82 variants in ncRNA_exonic, 5445 variants in ncRNA_intronic, 24 variants in splicing, 5 variants in upstream, and 1 variant in exonic splicing region. Among the other variant types, 910 variants were synonymous, and 1462 variants were non-synonymous. Moreover, 59 frameshift, 40 non-frameshift, 8 start-loss, 3 stop-gain, stop-loss variants were also found. However, 8 variants types remained unknown (Figure 4.1 B).

A.



B.

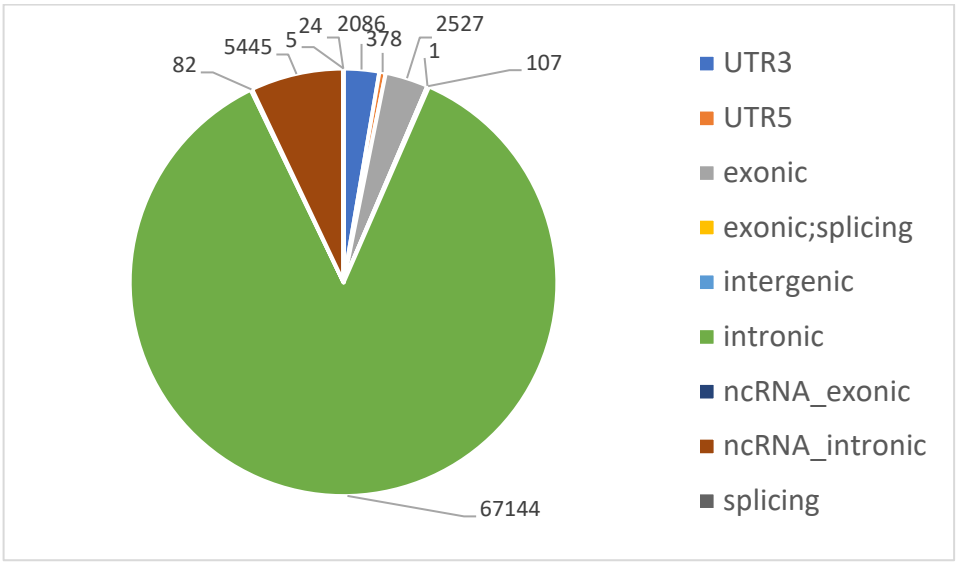


Figure 4.1 : Distribution of variant types: A. Variants in WES dataset. B. Variants in WGS datasets

We have constructed the **M**izoram **P**harmacogenomic **Var**iant **D**ata**b**ase (MPVARdb v.1) to have better access to each variant type. The relevant details such as allele frequency in the Mizo population and 1000g project can be searched using the gene name in the HUGO gene nomenclature system (HGNC). The database accepts the pharmacogenetic categorized as VIP in PharmGKB and can be searched in the database to get the details of the variant (Figure 4.2).

The result page provides pie charts for variant types found in WES (healthy exomes) and WGS (Indigen NE) datasets. It provides two pie charts from each dataset, the first pie charts provides the details of the distribution of synonymous, nonsynonymous variants and other variants, and the second pi-chart provides the details of variants based on location. In the database result page, fetched data from WES displays first in the form of two pie charts (Figure 4.3 A), PharmGKB classification (Figure 4.3 B), database total number of hits found against the search term, and variant in tabular format including the allele frequency from the WES datasets (Figure 4.3 C). The database result page displays the results in terms of the variant details in a detailed tabular format derived from WES as well as WGS data. Similarly, it also shows the two pie charts (Figure 4.4 A), database total number of hits

found against the search term, the variants in a tabular format including the allele frequency from the WGS datasets (Figure 4.4 B)

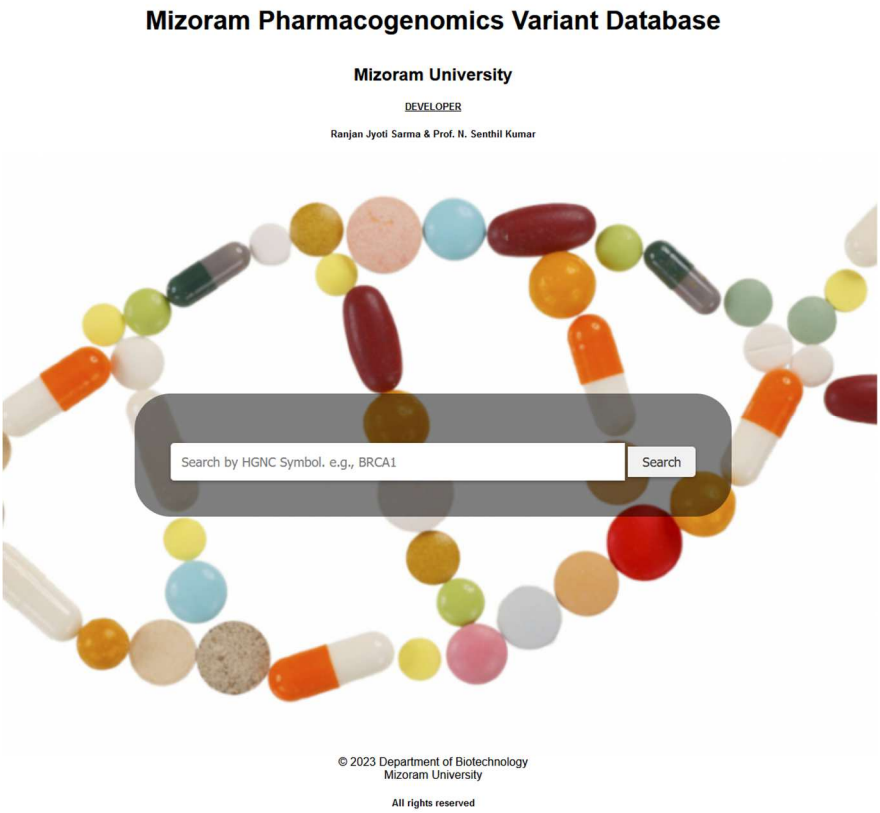
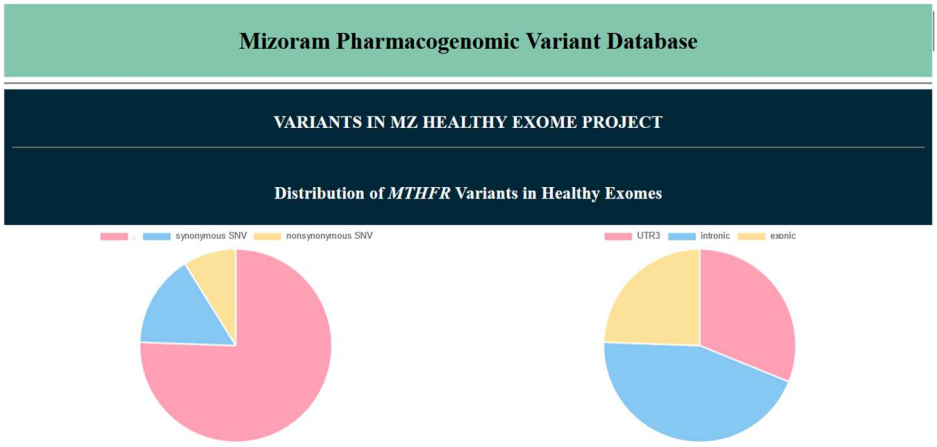


Figure 4.2 : The home page of the MPVARdb. The search bar accepts the Gene name in the HGNC symbol and scans it in the database.

A.



B

PharmGKB CLASSIFICATION

MTHFR

Methylenetetrahydrofolate reductase (MTHFR) plays a key role in folate metabolism and is implicated in the pharmacodynamics of several drugs. For more: https://www.pharmgkb.org/vip/PA166169429

PharmGKB Classification

VIP Tier 1

PharmGKB Classification

Enzyme

C

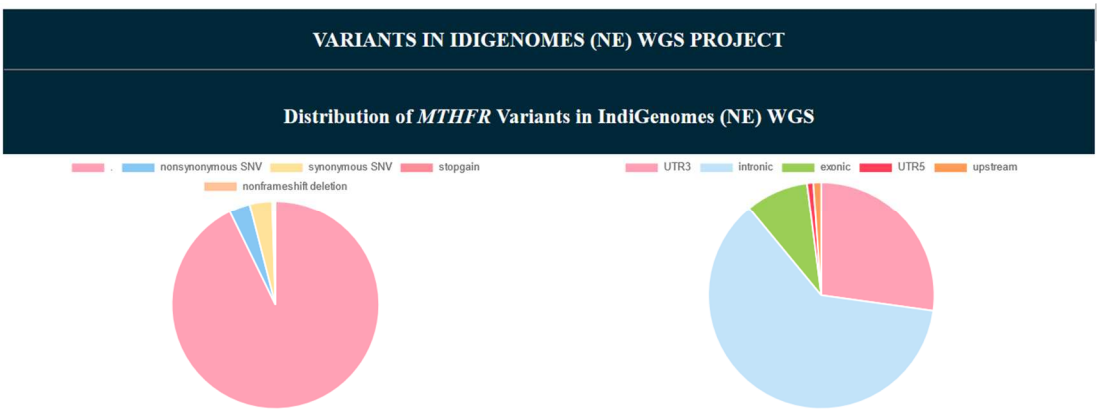
Total 45 variants for *MTHFR* found based on WES!

NOTE: The chromosomal positions are as per GRCh38 reference genome.

Sample	Chromosome	Start	End	Aleteration	HGNC	Func_refGene	dbSNP Entry	MZ_AF_ExP	AF_1000g_All	Clinical Significance
GI-11	chr1	11787277	11787277	0>GAA	MTHFR	UTR3	.	0.143	.	.
GI-11	chr1	11788822	11788822	A>G	MTHFR	UTR3	rs2077360	0.143	0.930511	.
GI-11	chr1	11791061	11791061	A>G	MTHFR	intronic	rs7518348	0.143	0.902157	.
GI-11	chr1	11794400	11794400	G>A	MTHFR	exonic	rs4846051	0.143	0.902955	Benign
GI-11	chr1	11794698	11794698	G>A	MTHFR	intronic	rs1994798	0.143	0.579273	.
GI-13	chr1	11787845	11787845	G>A	MTHFR	UTR3	.	0.13	.	.
GI-13	chr1	11790870	11790870	C>A	MTHFR	UTR3	rs1537516	0.093	0.111422	.
GI-13	chr1	11790946	11790946	C>T	MTHFR	UTR3	rs1537515	0.13	0.111022	.
GI-13	chr1	11792243	11792243	G>C	MTHFR	UTR3	rs2077360	0.148	0.930511	.
GI-13	chr1	11796321	11796321	C>T	MTHFR	exonic	rs2274976	1	0.0744808	"Benign,_other"
GI-13	chr1	11800063	11800063	G>A	MTHFR	intronic	rs3818762	0.111	0.246006	.
GI-13	chr1	59907994	59907994	T>C	MTHFR	intronic	rs7518348	0.028	0.902157	.
GI-13	chr1	97450068	97450068	A>C	MTHFR	intronic	rs1476413	0.028	0.251198	.
GI-13	chr1	169520708	169520708	A>G	MTHFR	exonic	rs4846051	0.083	0.902955	Benign
GI-13	chr1	169529597	169529597	A>G	MTHFR	intronic	rs1994798	0.083	0.579273	.
GI-13	chr1	169550527	169550527	C>T	MTHFR	intronic	rs11121832	0.963	0.759185	.
GI-16	chr15	74751209	74751209	T>C	MTHFR	UTR3	.	0.074	.	.
GI-16	chr15	74751878	74751878	G>C	MTHFR	UTR3	rs2077360	0.074	0.930511	.
GI-16	chr17	39715893	39715893	G>C	MTHFR	exonic	rs797015332	0.5	.	.

Figure 4.3: MPVardb result page: A. Pi chart of variant types from WES dataset, B. Variant classification in PharmGKB, C. Variant details extracted from WES dataset

A



B

Total 400 variants for *MTHFR* found based on WGS!

NOTE: The chromosomal positions are as per GRCh38 reference genome.

Chromosome	Start	End	Aleteration	HGNC	Func_refGene	dbSNP Entry	MZ_AF_wgsP	AF_1000g_All	Clinical Significance
chr1	11785774	11785774	G>A	MTHFR	UTR3	rs563513836	1	0.000399361	.
chr1	11785823	11785823	C>-	MTHFR	UTR3	.	1	.	.
chr1	11785882	11785882	G>A	MTHFR	UTR3	rs886045166	1	.	.
chr1	11785889	11785889	C>T	MTHFR	UTR3	rs12023469	0.913978	0.0457268	.
chr1	11785908	11785908	T>A	MTHFR	UTR3	rs546580031	1	0.000199681	.
chr1	11785922	11785922	G>A	MTHFR	UTR3	rs557701148	1	.	.
chr1	11785948	11785948	G>C	MTHFR	UTR3	.	1	.	.
chr1	11785953	11785953	G>A	MTHFR	UTR3	rs1016519896	1	.	.
chr1	11786025	11786025	G>A	MTHFR	UTR3	.	1	.	.
chr1	11786035	11786035	->T	MTHFR	UTR3	rs55740775	0	0.951278	Benign
chr1	11786093	11786093	A>G	MTHFR	UTR3	rs925919532	1	.	.
chr1	11786147	11786147	C>A	MTHFR	UTR3	.	1	.	.
chr1	11786195	11786195	G>A	MTHFR	UTR3	rs4846048	0.16129	0.70647	.
chr1	11786208	11786208	G>A	MTHFR	UTR3	.	0.994624	.	.
chr1	11786213	11786213	T>C	MTHFR	UTR3	rs1057624	0.994624	0.00998403	.
chr1	11786295	11786295	A>C	MTHFR	UTR3	.	1	.	.
chr1	11786378	11786378	C>T	MTHFR	UTR3	rs534760603	1	0.000599042	.
chr1	11786390	11786390	G>A	MTHFR	UTR3	rs4845884	0	0.933307	.
chr1	11786427	11786427	->T	MTHFR	UTR3	.	0.989247	.	.
chr1	11786427	11786427	T>-	MTHFR	UTR3	rs112469003	0.989247	.	.
chr1	11786454	11786457	CAGA>-	MTHFR	UTR3	.	1	.	.
chr1	11786488	11786488	T>C	MTHFR	UTR3	.	1	.	.

Figure 4.4: MPVardb result page: A. Pi chart of variant types from WGS dataset, B. Variant details extracted from WGS dataset

4.2 Clinically Actionable PGx Variants in Healthy Population

The clinically actionable (CA) PGx variants were screened using PharmCat annotation tools that integrated PharmaGKB-DPWG and CPIC guidelines. The annotation was performed using PharmCAT in healthy exomes (n=27) to study the prevalence of CA PGx variants with associated star alleles that have decreased function, intermediate function, and variable function. Furthermore, *DPYD* star alleles were also investigated in the datasets.

There were multiple important CA PGx variants were observed in the healthy exome datasets. A variant Chr4:88131171 G>T (rs2231142) in *ABCG2* with allele frequency (AF) in the healthy exome and Indigen (NE datasets) projects were found to be 0.259 and 0.811, respectively. The variant is responsible for decreased function (DF) for the drugs allopurinol and rosuvastatin. There was no related star allele detected in the variant, however, PharmaGKB-DPWG's clinical recommendations were available.

Another variant Chr19:40991369 C>T (rs8192709) was detected *CYP2B6* gene responsible for intermediate function (IF) for Efavirenz, Sertraline. The reference allele C is *1, while the altered allele T is *10 for the variant can also be represented as CYP2B6*10. The variant AF in healthy exome and IndiGen NE datasets were found to be 0.111 and 0.913, respectively. Another important gene *TPMT* variant Chr6:18130687 T>C (rs1142345) or *TPMT*3A* was detected in one donor (AF= 0.01). The variant AF in the Indigen NE dataset was found to be 0.961. The variant is an intermediate metabolizer for thiopurine-based drugs Azathioprine, Mercaptopurine, and Thioguanine and both the PharmGKB-DPWG and CIPIC provided the recommendation with clinical guidelines.

Another variant ChrX:154534495 C>T in *G6PD* (rs137852314) or *G6PD*-Mahidol variant was detected in the healthy exome that is responsible for variable function in the metabolism of drugs like Aspirin, Chloramphenicol, Chloroquinon, Norfloxacin, exfloxacin, Quinine and many more. The AF in healthy exome and NE IndiGen dataset was found in 0.05 and 0.989, respectively.

The variant Chr2:233760498 G>A (rs4148323) or *UGT1A1*6* (the altered allele denoted by *6) variant with AF 0.13 was detected in the healthy exome, affects effect the metabolism of Atazanavir, Irinotecan. The variant AF in indigen NE datasets was found to be 0.779.

The variant Chr12:21176804 A>G (rs2306283) or *SLCO1B1*14* potentially leads to a decrease in the metabolism of statins. The variants AF in healthy exome and IndiGen NE dataset were found to be 0.722 and 0.354, respectively. Clinical guidelines for the variant have been provided by CIPIC for the variant *SLCO1B1*14* (Table 4.1).

Two *DPYD* variants were detected by PharmCAT with normal function annotation. The variant Chr4:88131171 G>T (rs1801159) or *DPYD*5* with AF in healthy exome and NE Indigen datasets were 0.255 and 0.725, respectively. Another variant Chr19:40991369 C>T (rs17376848) in *DPYD* was observed with AF in healthy exome and IndiGen NE datasets as 0.111 and 0.924, respectively. No star allele assigned to the variant rs17376848 yet. There are no clinical guidelines provided by PharmGKB-DPWG and CIPIC (Table: 4.2).

Table 4.1: Variants of clinical importance annotated using PharmCAT in WES data of healthy exome data.

dbSNP/Gene	Variant	Star Allele	MZ_AF	IndiGen AF	PharmCAT	Prescribing Recommendation		Drugs
						PharmGKB-DPWG Guidelines	CIPIC Guidelines	
rs2231142 *ABCG2*	Chr4:88131171 G>T	-	0.259	0.811	DF	Yes	No	Allopurinol, Rosuvastatin
rs8192709 *CYP2B6*	Chr19:40991369 C>T	*1/*10	0.111	0.913	IF	No	No	Efavirenz, Sertraline
rs1142345 *TPMT*	Chr6:18130687 T>C	*1/*3A	0.01	0.961	IF	Yes	Yes	Azathioprine, Mercaptopurine, Thioguanine
rs137852314 *G6PD*	ChrX:154534495 C>T	B/Mahidol	0.05	0.989	V	No	Yes	Aspirin, Chloramphenicol, Chloroquine, Norfloxacin, Ofloxacin, Quinine, And many other important drugs.
rs4148323 *UGT1A1*	Chr2:233760498 G>A	*1/*6	0.13	0.779	IF	No	Yes	Atazanavir, Irinotecan
rs2306283 *SLCO1B1*	Chr12:21176804 A>G	*1/*14	0.722	0.354	DF	No	Yes	Atorvastatin, Fluvastatin, Lovastatin, Pitavastatin, Pravastatin, Rosuvastatin, Simvastatin

CPIC- The Clinical Pharmacogenetics Implementation Consortium; DF- Decreased Function; IF-Intermediate Function; V-Variable

Table 4.2: *DPYD* variant annotated using PharmCAT in WES data from healthy exome data.

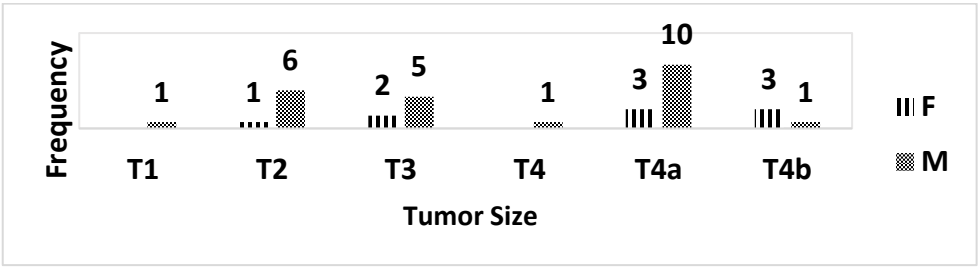
dbSNP/Gene	Variant	Star Allele	AF in Mizoram	PharmCAT	Prescribing Recommendation		Drugs
					PharmGKB-DPWG Guidelines	CIPIC Guidelines	
rs1801159 *DPYD*	Chr1: 97515839T>C	*5	0.259	NF	No	No	• 5FU, • Capecitabine • Tegafure
rs17376848 *DPYD*	Chr19:40991369 C>T	-	0.111	NF	No	No	

CPIC- The Clinical Pharmacogenetics Implementation Consortium

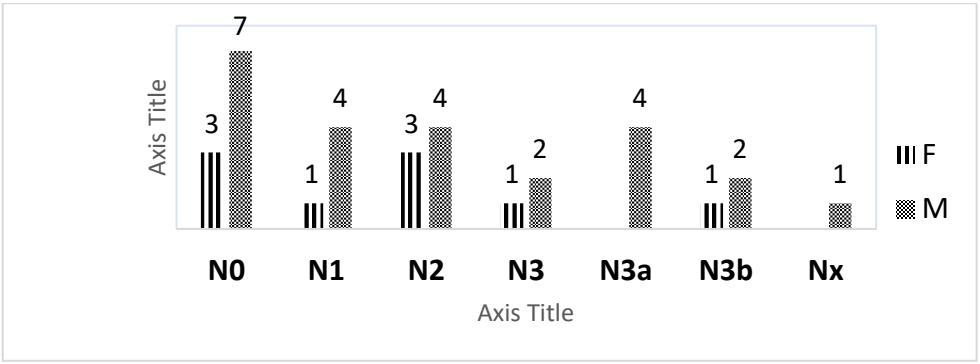
4.3.1 TNM Staging of the patients with ADR

The tumor size (T) data was available for 33 out of the 37 follow-up patients and it was observed that a greater number of the patients were diagnosed with T_{4a} (39.39%) followed by T_3 (21.21%) and T_2 (21.21%) **(Figure 4.5 A)**. Lymph nodes followed by numbers signify the presence of cancer in the nearby lymph nodes. Ten patients (33.33%) did not have cancer in lymph nodes (N_0). However, one patient was diagnosed in N_{4a} stage **(Figure 4.5 B)**. The metastasis occurred in 6 patients; however, it was unknown or could not be detected **(Figure 4.5 C)**.

A



B



C

Figure 4.5: Comparison of Male and Female proportions based on: A. Tumor Size, B. Tumor Nodes, and C. Unknown metastasis (M_x)

4.3.2 Adverse drug reaction in Gastric cancer patients

We have collected follow-up data from 37 gastric cancer patients showing ADR after chemotherapy. We have documented almost 29 ADRs of different grades. The numbers of female and male patients were 9 (24.32%) and 28 (75.68%), respectively (Figure 4.6).

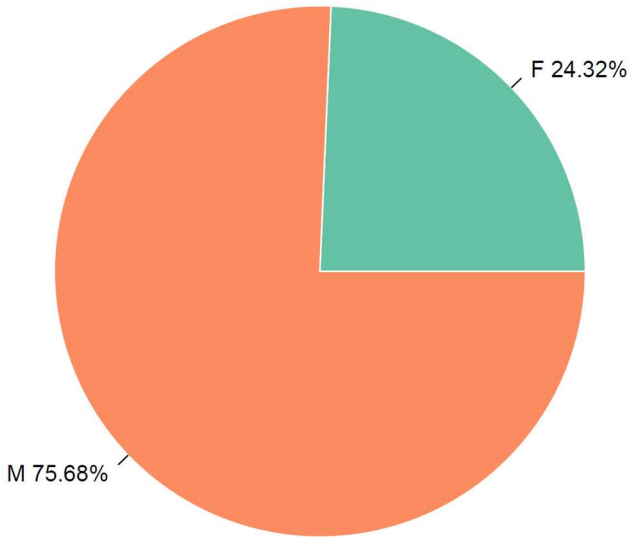


Figure 4.6: Proportion of Male (M) and Female (F) Gastric cancer patient showing ADR

4.3.3 Types of ADRs

There were prominent ADRs of different grades observed after chemotherapy: Loss of Appetite (26 Patients), Fatigue (21 Patients), Nausea (23 patients), Drowsiness (18 patients), Dermatological Adverse Reaction (DAR) (11 patients), Vomiting (8 patients) and Alopecia characterised by hair loss (8 Patients) along with other ADRs. Moreover, neutropenia in 3, leukopenia in 2 and low haemoglobin in 1 patient were revealed in CBC profiling after chemotherapy **(Figure 4.7)**. However, the ADRs were observed with other ADRs concurrently in the same patients. Loss of appetite was observed in 26 patients along with other types of ADRs. Loss of Appetite with Fatigue, Nausea, Drowsiness was observed in 3 patients and with Nausea, Drowsiness in another 3 patients **(Table 4.3)**.

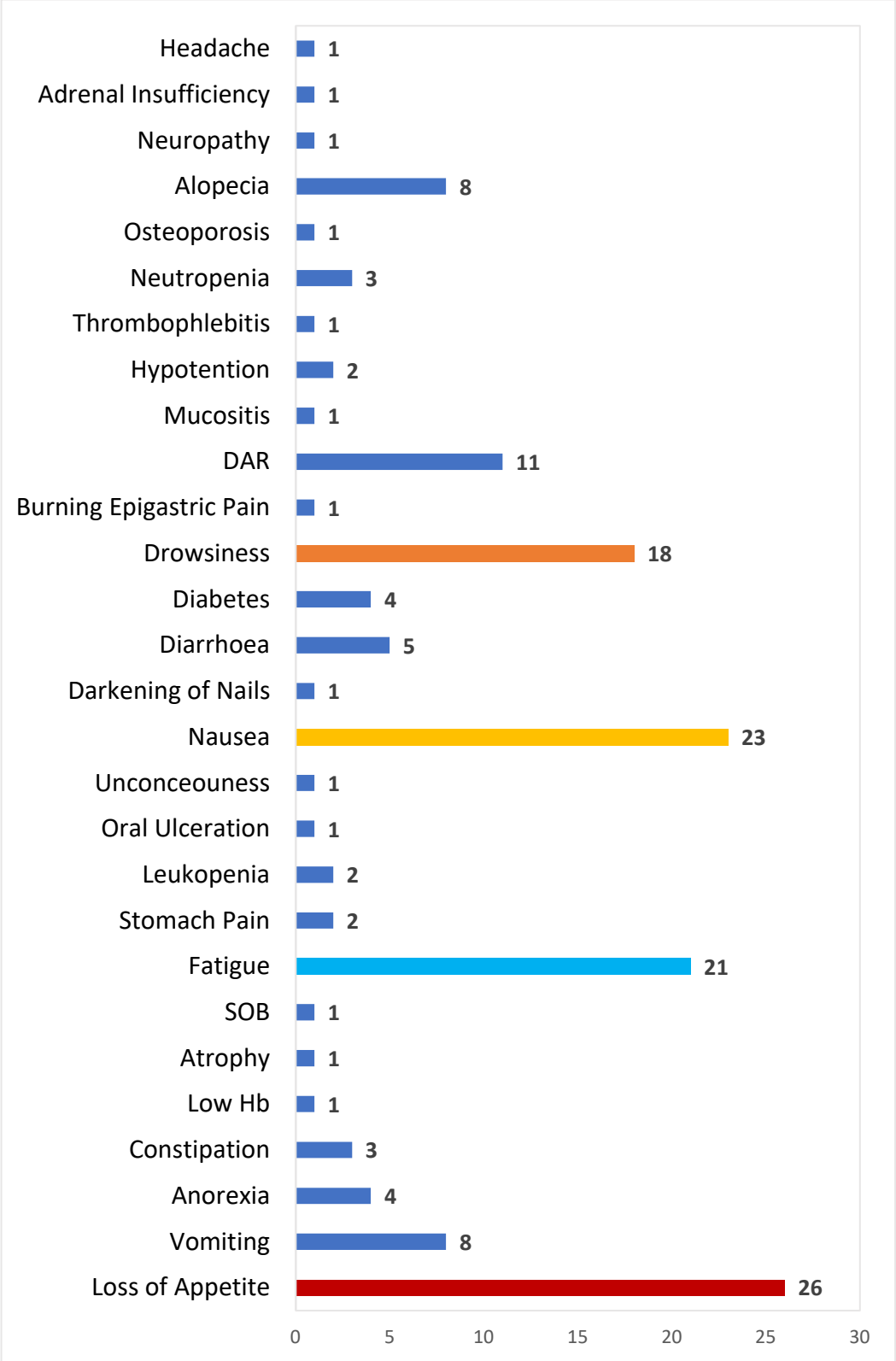


Figure 4.7: Different types of ADRs observed in GC patients after Chemotherapy (DAR- Dermatological Adverse Reactions; SOB- Shortness of Breath)

Table 4.3: ADR with other comorbidities

ADRs	ADRs as Comorbidities	Number of Patients
Loss of Appetite	Anorexia, Diabetes	1
	Anorexia, Fatigue, Nausea, Drowsiness	1
	Anorexia, Fatigue, Nausea, Drowsiness, Alopecia, Neuropathy, Adrenal Insufficiency	1
	Constipation, Fatigue, Nausea, Drowsiness, DAR, Burning Epigastric Pain, Headache	1
	Atrophy, Nausea, Vomiting, Fatigue, Leukopenia, Diabetes, Drowsiness, DAR	1
	Fatigue	1
	Fatigue, Stomach Pain, Diabetes	1
	Fatigue, Leukopenia, Hypotension	1
	Fatigue, Nausea, Leukopenia, Vomiting, Diabetes, Drowsiness, DAR	1
	Fatigue, Oral Ulceration, Diarrhoea, Drowsiness,	1
	Fatigue, Unconsciousness	1
	Fatigue, Nausea	1
	Fatigue, Nausea, Drowsiness	3
	Fatigue, Neutropenia	1
	Fatigue, Nausea, Vomiting, Drowsiness, DAR, Thrombophlebitis, Alopecia	1
	Fatigue, Drowsiness	1
	Fatigue, Nausea, Vomiting, Diabetes, Drowsiness, DAR, Hypotension	1
	Nausea, Drowsiness	3
	Nausea, Diarrhoea, DAR,	1
	Diarrhoea, DAR	1
	Drowsiness	1
	Anorexia, Fatigue, Nausea, Drowsiness	1
Anorexia	SOB, Fatigue, Vomiting	1
Constipation		1
Constipation	Fatigue, Stomach Pain, Oedema, Nausea, Vomiting,	1
Low HB	Mucositis, Neutropenia,	1
Fatigue	Nausea, Drowsiness	1
Nausea	Alopecia	1
	Vomiting, DAR,	1
	Drowsiness	1
	Vomiting, DAR,	1
Drowsiness		1
Osteoporosis		1
	Total	37

4.4.1 Adjuvant Therapy in Gastric Cancer

Out of 59 patients, 37 patients were followed up for adverse drug reactions and received chemotherapy after surgery. However, exome sequencing data as well as patient followed-up data could be performed for 25 patients. 5FU was administered to 23 patients (19 patients by intravenous and 4 patients by oral route). One patient was given paclitaxel and another patient was given Oxaliplatin with Leucovorin, Cisplatin, and Docetaxel (Capecitabine) (Table 4.4).

Table 4.4: Chemotherapy regimen used in gastric cancer patients from Mizoram.

Chemo	Total Count	Samples
Capicitabin	4	gc-106, gc-139, gc-141, gc-41
5-Fluorouracil +Cisplatin + Oxaliplatin	1	gc-40
5-Fluorouracil with Leukovorin + Oxaliplatin	18	gc-130, gc-131, gc-135, gc-17, gc-2, gc-35, gc-42, gc-58, gc-59, gc-64, gc-75, gc-76, gc-79, gc-81, gc-83, gc-85, gc-9, gc-98
Oxaliplatin with Leukovorin+Cisplatin+Docetaxal	1	gc-86
Paclitaxel	1	gc-18

4.4.2 Fluorouracil PK/PD Pathway gene variants in GC patients

The follow-up data suggest that the majority of GC patients receive FLO as adjuvant chemotherapy. The genes *DPYD, DPYS,* and *UPB1* play major role in metabolism of 5-Fluorouracil. However, the drug oxaliplatin undergoes non-enzymatic transformation and therefore the 5-Fluorouracil (5FU) PK/PD genes variants were investigated in all the 59 GC patients. A variant in Chr1(GRCh38):97515839T>C or I543V (db) of *DPYD* (dbSNP: rs1801159) associated for DPD deficiency or toxicity was found in 35% of patients. Moreover, the variant Chr8(GRCh38):104466705G>A or (F72F) (dbSNP: rs2298840) of *DPYS* associated with Dihydropyrimidinase deficiency was found in 45.76% of patients. Another 5'UTR region variant Chr22(GRCh38):24495387A>T or (NM 016327:c.17A>T) (dbSNP: rs2070475) in *UPB1* associated with deficiency of beta ureidopropionase was found in 45.15% patients. The variant in *TYMS* gene chr18:662247A>G or (E127E) (dbSNP: rs3786362) was detected in 62.06% of patients **(Table 4.5).**

Table 4.5: PK/PD Pathway gene variants among Gastric cancer patient (n=59):

Variant	variant ID	Amino Acid Change	Clinvar Status	Number of Patients	% of positive samples for the variant
DPYD					
Chr1: 97515839T>C	**rs1801159**	**NM_000110:exon13:c.A1627G:p.I543V**	• **DPD deficiency** • **Capecitabine response – toxicity** • **Fluorouracil response - Toxicity**	**21**	**35**
Chr1:97305279G>A	rs112766203	NM_000110:exon18:c.C2279T:p.T760I	• DPD deficiency	1	1.6
Chr1:97305364C>T	rs1801160	NM_000110:exon18:c.G2194A:p.V732I	• DPD deficiency	1	1.6
Chr1: 97883329A>G	rs1801265	NM_001160301:exon2:c.T85C:p.C29R	• DPD deficiency	7	11.8
Chr1:97450068A>G	rs17376848	NM_000110:exon14:c.T1896C:p.F632F	• DPD deficiency	6	10.16
DPYS					
Chr8:104466705G>A	**rs2298840**	**NM_001385:exon1:c.C216T:p.F72F**	• **Dihydropyrimidinase deficiency**	**27**	**45.76**
Chr8: rs36027551G>A	rs36027551	NM_001385:exon3:c.C541T:p.R181W		5	8.47
UPB1					
Chr22: 24495387A>T	**rs2070475**	**NM_016327:c.-17A>T**	• **Deficiency of beta**	**29**	**49.15**
Chr22: 24525761G>A	rs35916595	NM_016327:exon10:c.G1122A:p.K374K	**ureidopropionase**	2	3..8
TYMS					
chr18:662247A>G	**rs3786362**	**NM_001071:exon3:c.A381G:p.E127E**	**Not reported**	**18**	**30.5**

Table 4.6: *DPYD* variants in gastric cancer patients followed up post chemo therapy (n=23).

Sample	Chemo	Nucleotide Change	Variant Type	Amino Acids Change	dbSNP	Clinicial Outcome	Clinical Significance	Vital Status	1000g AF	Zygos
gc-2	FLO							Deceased		
gc-35	FLO							Deceased		
gc-41	Capicitabin					DPD deficiency , Capecitabine response – toxicity, Fluorouracil response - Toxicity		Deceased		
gc-59	FLO	T>C	Nonsynonymous	NM_000110:exon13:c.A1627G:p.I543V	rs1801159		Benign	Deceased	0.184904	Het
gc -75	FLO							Alive		
gc -76	FLO							Deceased		
gc -85	FLO							Alive		
gc -130	FLO	G>A	Nonsynonymous	NM_000110:exon18:c.C2279T:p.T760I	rs112766203	DPD deficiency	VUS	Deceased	0.001198	Het
gc -135	FLO	C>T	Nonsynonymous	NM_000110:exon18:c.G2194A:p.V732I	rs1801160	DPD deficiency	Benign/Likely Benign	Alive	0.0439297	Het
gc -40	FOCis	A>G	Nonsynonymous	NM_000110:exon2:c.T85C:p.C29R	rs1801265	DPD deficiency	Benign	Alive	NA	Het
gc -141	Capicitabin									
gc 64	FLO	A>G	Synonymous	NM_000110:exon14:c.T1896C:p.F632F	rs17376848	DPD deficiency	Benign	Alive	0.052117	Hon
gc -85	FLO									Het

Table 4.7: ADRs in the patients with the nonsynonymous variant NM_000110:exon13:c.A1627G:p.I543V (rs1801159) in *DPYD*

Sample		Vital Status	ADRs (Grade)
gc-2	FLO	Deceased	Osteoporosis (3)
gc-35	FLO	Deceased	LoA(2), Vomiting (1), DAR (1), Alopecia (2)
gc-41	Capicitabin	Deceased	Constipation (3)
gc-59	FLO	Deceased	LoA (2), Fatigue (2), Nausea (2), Vomiting (3), Drowsiness (1), DAR (2), Thrombocytopenia (3), Alopecia (1)
gc-75	FLO	Alive	LoA (3), Fatigue (2), Nausea (3), Vomiting (2), Diabetes (1), Drowsiness (2), DAR (3), Hypotension (3)
gc-76	FLO	Deceased	LoA (1), Fatigue (1), Nausea (1), Drowsiness (1)
gc-85	FLO	Alive	LoA (2), Fatigue (2), Nausea (2)

Table 4.8: *DPYS* variants in gastric cancer patients followed up post chemo therapy (n=23).

Sample	Chemo	Nucleotide Change	Variant Type	Amino Acids Change	dbSNP	Clinical Outcome	Clinical Significance	Vital Status	1000g AF
gc -2	FLO							Deceased	
gc -17	FLO							Deceased	
gc -35	FLO							Deceased	
gc -41	Capicitabin							Deceased	
gc -42	FLO							Deceased	
gc -58	FLO	G>A	synonymous	NM_001385:exon1:c.C216T:p.F72F	rs2298840	Dihydropyrimidinase deficiency	Benign	Deceased	0.231629
gc -64	FLO							Alive	
gc -85	FLO							Alive	
gc -86	OLCD							Alive	
gc -106	Capicitabin							Deceased	
gc -131	FLO							Alive	
gc -35	FLO	G>A	synonymous	NM_001385:exon3:c.C541T:p.R181W	rs36027551	Dihydropyrimidinase deficiency	Benign	Deceased	

dbSNP- SNP database identifier; 1000g AF- Allele Frequency (AF) in 1000 Genome datasets

Table 4.9: *UPB1* variants in gastric cancer patients followed up post chemo therapy (n=23).

Sample	Cheno	Nucleotide Change	Variant Type	cDNA Change/Amino Acids Change	dbSNP	Clinicial Outcome	Clinical Significance	Vital Status	1000g AF
gc -17	FLO							Deceased	
gc -58	FLO							Deceased	
gc -79	FLO	A>T	UTR5	NM_016327:c.-17A>T	rs2070475	Deficiency of beta ureidopropionase	Benign	Deceased	0.15595
gc-83	FLO							Alive	
gc -98	FLO							Alive	
gc -79	FLO	G>A	Synonymous	NM_016327:exon10:c.G1122A:p.K374K	rs35916595	Deficiency of beta ureidopropionase	Benign	Deceased	0.00419329

dbSNP- SNP database identifier; 1000g AF- Allele Frequency (AF) in 1000 Genome datasets

Table 4.10: *TYMS* variants in gastric cancer patients followed up post chemo therapy (n=23).

Sample	Chemo	Nucleotide Change	Variant Type	Amino Acids Change	dbSNP	Clinical Outcome	Clinical Significance	Vital Status	1000g AF	Zygosity
gc -2	FLO	A>G	synonymous	NM_001071:exon3:c.A381G:p.E127E	rs3786362	Not reported in database	VUS	Deceased	0.0623	Het
gc -58	FLO							Deceased		Hom
gc -75	FLO							Alive		Het
gc -76	FLO							Alive		Het
gc -85	FLO							Alive		Hom
gc -98	FLO							Alive		Het
gc -130	FLO							Deceased		Het
gc -131	FLO							Alive		Het
gc -139	Capicitabin							Alive		Het
gc -141	Capicitabin							Alive		Het

4.4.3 Variants in genes involved in the Pharmacokinetic pathway of 5FU: *DPYD* variants and observed ADRs

Exome sequencing revealed that 12 patients (52.17%) out of 23 patients received 5-Fluorouracil had *DPYD* variants responsible for DPD Deficiency. A variant NM_000110:exon18:c.C2279T:p.T760I (rs112766203) was observed in a patient with advanced gastric cancer and received adjuvant chemotherapy of 5-Fluorouracil with leucovorin and Oxaliplatin. The allele frequency (AF) of the variant in the 1000g (1000 Genome Project) study was found to be 0.001198, categorized as a rare variant (Table 4.6). In the patient, there were various ADRs of different grades namely: Loss of Appetite (Grade 1), Nausea (Grade 1), Darkening of Nails (Grade 2), severe neutropenia, and Alopecia (Grade 1). The patient survived 512 days (17 months and 10 days) from the date of detection of cancer. The variant was predicted to be damaging by SIFT, Polyphen, and Mutation taster.

Another nonsynonymous variant NM_000110:exon18:c.G2194A:p.V732I (rs1801160) was observed in one patient. The variant AF was found to be 0.0439297. The variant was categorized as benign or likely benign in the ClinVar database. Similarly a nonsynonymous variant NM_000110:exon2:c.T85C:p.C29R (rs1801265) was observed in two patients. Both variants were reported to be associated with DPD deficiency and found in heterozygous conditions. One Synonymous variant NM_000110:exon14:c.T1896C:p.F632F (rs17376848) was detected in two patients. The allele frequency of the variant type was found to be 0.052117. The variant was classified as benign in the ClinVar database. All three patient is surviving up to the last follow-up conducted on 4 May 2023.

Another nonsynonymous variant type responsible for DPD deficiency NM_000110:exon13:c.A1627G:p.I543V (rs1801159) was detected in seven (~30%) out of 23 patients who received 5-Fluorouracil . This variant can also be represented as *DPYD*5*. It was found that five of the seven patients were deceased and two patients are still surviving. Four out of the five deceased patients received 5-Fluorouracil with leucovorin and Oxaliplatin (FLO) (6 to 8 Cycles) and one patient received oral capecitabine (4 Cycles). Survival data was received from only three patients out of five diseased patients. One patient survived for less than a year (red), one patient survived

for more than a year but less than two years (green) and one patient survived for more than two years but less than three years (Orange) [Figure 4.8].

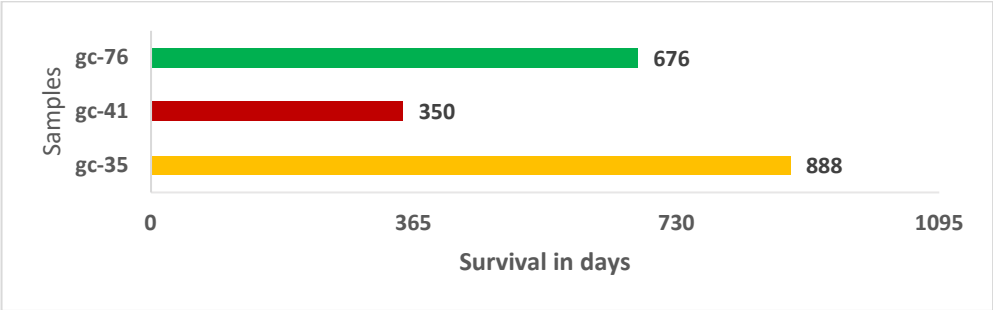


Figure 4.8: Survival of the patient with *DPYD* variant (rs1801159).

A total of 4 Nonsynonymous variants were observed in only the *DPYD* gene among the PK/PD pathway of 5FU. Among the patient positive for rs1801159, four patients experienced Loss of Appetite (LoA) with grade 2 or higher and one patient shared grade 1 LoA. Three patients had Dermatological Adverse reactions (DAR) with grades 1,2 and 3. Grade 2 Fatigue was observed in 3 patients, grade 1 in one patient. Grade 2 alopecia was observed in two patients. Nausea was observed as grade 2 in two patients, grade 3 in one patient, and grade 1 in one patient. Mild diabetes was observed in one patient and grade 3 thrombocytopenia was characterized by low blood platelet levels. Constipation was observed in only one patient. Mild to moderate drowsiness was observed in 3 patients. Similarly, mild to moderate vomiting was observed in two patients, and severe vomiting in one patient (Table 4.7).

4.4.4 *DPYS* Variants and observed ADRs

Two types of *DPYS* gene variants were observed in 11 patients out of 23 patients where a particular variant NM_001385:exon1:c.C216T:p.F72F (rs2298840) was observed in 11 patients (47.82%) alone. Out of the 11 patients, 7 patients were deceased and 4 patients were surviving up to the latest follow-up. The patient with the variant rs36027551 was also deceased and variant was not observed in the 1000g annotation datasets. The AF of the variants rs2298840 in 1000g datasets was 0.231629 (Table 4.8). One patient (gc-35) had the two *DPYS* variants rs2298840 and NM_001385:exon3:c.C541T:p.R181W (rs36027551) along with one *DPYD* variant rs1801159 that are associated with DPD deficiency and Dihydropyrimidinase

deficiency found to be diseased (Table 4.6, Table 4.8). It was observed that the majority of the deceased patients (n=6) survived less than 5 years (1825 days).

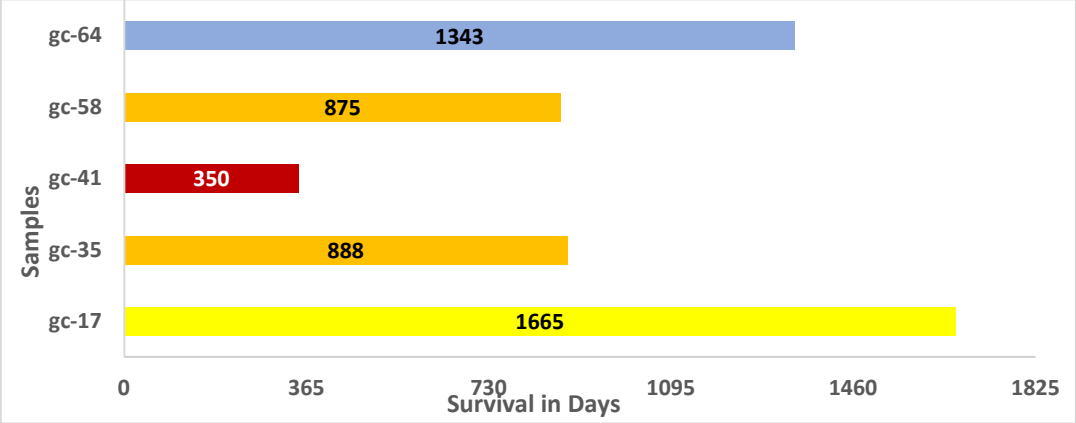


Figure 4.9: Survival (in days) of the patients with *DPYS* variants. Two patients survived less than a year (Red), two patients survived more than a year but less than 3 years (Orange), one patient more than 3 years but less than 4 years (light blue), one patient more than 4 years but less 5 years (Yellow), One patient survived for more than 5 years but less than 8 years (green).

4.4.5 *UPB1* variants and observed ADRs

Surprisingly, only one exonic variant (synonymous) NM_016327:exon10:c.G1122A:p.K374K (rs35916595) was observed in the *UPB1* gene in the heterozygous condition in one patient (gc-79). The variant is reported for deficiency in bete ureidopropionase, an enzyme that converts luoro-beta-ureidopropionate to fluoro-beta-alanine (FBAL). The AF of the variant in 1000g datasets was found to be 0.00419329. Another variant in 5'-UTR NM_016327:c.-17A>T (rs2070475) was also in the same patient (Table 4.9). The patient experienced grade 3 mucositis with neutropenia and survived for 548 days from the date of detection of the disease. Moreover, another 2 patients gc17 and gc-58 carrying the variant rs2070475 survived for 1665 days and 875 days, respectively. Different types of adverse reactions were observed in these patients including severe forms of anorexia, SOB, Diarrhoea and vomiting in gc17, mild forms of LoA, Anorexia, and diabetes in gc-58 and low haemoglobin (grade 2), mucositis (grade 3), and neutropenia (grade 3) in gc-79. Overall, 3 patients were found to be deceased out of 6 patients with

UPB1 variants and 3 patients are still surviving. One of the alive patients (gc-83) experienced LoA (grade 3), fatigue (grade 2), nausea (grade 3), drowsiness (grade 3), DAR (grade 3), Alopecia (grade 3). The other alive patient experienced LoA (grade 2), Diarrhoea (grade 2), DAR (grade 3) and Alopecia (grade 2) ADRs.

4.4.6 ADRs in patients with variants in more than one 5FU PK/PD pathway genes

Grade 3 osteoporosis as ADR was observed in one patient (gc-2) and Loss of Appetite (grade 2), Fatigue (grade 2), and Nausea (grade 2) in another patient with variants rs1801159 in *DPYD*, rs2298840 in *DPYS*, and rs3786362 in *TYMS*. Loss of Appetite (grade 2), Vomiting (grade 1), Dermatological Adverse Reaction (Grade 1), and Alopecia (grade 2) were observed in a patient (gc-35) with two synonymous variants in DPYS (rs36027551 and rs2298840) and nonsynonymous variants in *DPYD* (rs1801159). Constipation was observed in one patient with rs1801159 in *DPYD* and rs2298840 in *DPYS.* Similarly, two patients (gc 75 and gc76), rs3786362 in *TYMS* and rs1801159 in *DPYD* had experienced more than three ADRs. LoA (grade 3), Nausea (grade 3), Vomiting (grade 2), Diabetes (grade 1), Fatigue (grade 2) Drowsiness (grade 2), DAR (grade 3), Hypotension (grade 3) in one patient (gc-75), and LoA (grade 1), Fatigue (grade 1), Nausea (grade 1), Drowsiness (grade 1) were observed in another patient (gc-76).

4.4.7 eQTL ANALYSIS for the PK/PD variants

eQTL analysis of the variants of *DPYD*, *DPYS*, *UPB1* and *TYMS* were performed in GTEx data. The gene expression associated with their variants can explanied by net effect size (NES) and Beta distribution-adjusted empirical p-values calculated by FastQTL. All the siginificant cis-eQTL were considered in this study (Table 4.11).

No significant eQTL found for the variants rs1801159, rs112766203, rs17376848 of *DPYD*, rs2298840, rs36027551 variants of *DPYS* and rs3786362 of *TYMS*.

Significant eQTL (p-value <0.01) was found for the variants rs1801160 and rs1801265 of *DPYD*. The variant rs1801160 caused slight upregulation *DPYD* in Testis (NES: 0.57, p-value: 4.6e-9) and down regulation in Skin (NES:-0.28, p-value: 0.00012). Another variant, rs1801265 caused similar upregulation of *DPYD* in Esophagus mucosa (NES: 0.16, p-value: 3.9e-8) and Skin (NES: 0.16, p-value: 0.0000096).

Total eight significant eTLS found for the variant rs2070475 of *UPB1*. The variant causes upregulation of *UPB1* in subcutaneous adipose tissue (NES: .030, p-value: 2.1e-9), Tibial nerve (4.5e-8), brain cotex (NES: 0.57, p-value: 6.5e-7), Thyroid (NES: 0.031, p-value: 6.9e-7), Whole blood (NES:0.21, p-value: 2.4e-6), nucleus accumbens of brain (NES: 0.44, p-value: 9.8e-6), Arota (NES: 0.31, p-value: 0.000012) and Putamen of Brain (NES: 0.46, p-value: 0.000024).

4.4.8 Survival Analysis With Gene Expression Data

eQTL analysis revealed the upregulation of *DPYD* and *UPB1* genes in the GTEx datasets due to certain variants. Therefore the gene expression and the overall survival (OS) probability were tested in KM-Plotter (https://kmplot.com) using gastric cancer datasets curated from Gene Expression Omnibus (GEO) datasets. A total of 157 patients who received 5FU-based adjuvant therapy were included in the test. The test was performed based on 60 months of follow-up records and was not restricted to a particular stage of cancer.

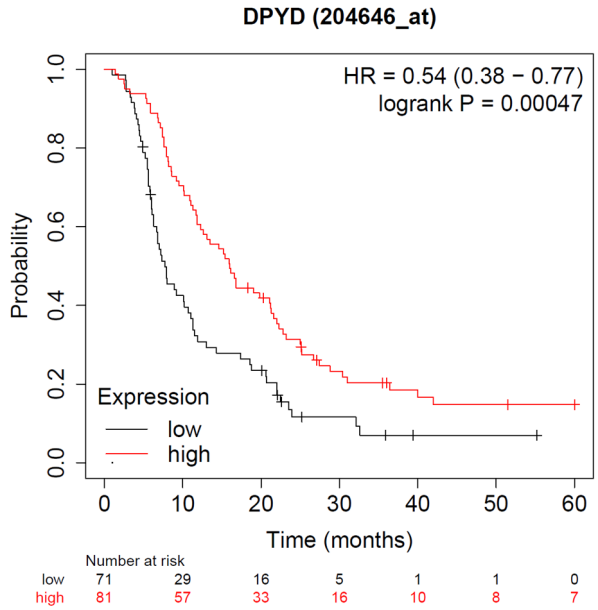
It was observed that *DPYD* upregulation was significantly associated with survival probability more than downregulation of the the gene (p-value: 0.00047). The hazard ratio (HR) in *DPYD* expression vs. survival was found to be 0.54, which suggests a low risk. On the other hand, *UPB1* upregulation was found to be associated with poor survival than downregulation but the association was found to be insignificant (p-value: 0.057). However, HR 1.47 *UPB1* expression vs. survival suggests a high risk of survival.

Table 4.11: Significant cis-eQTL of the PK/PD variants

Variant	dbSNP ID	p-value	NES	Tissue
DPYD				
Chr1: 97515839T>C	rs1801159	No significant eQTL found		
Chr1:97305279G>A	rs112766203			
Chr1:97305364C>T	rs1801160	4.6e-9	0.57	Testis
		0.00012	-0.28	Lung
Chr1: 97883329A>G	rs1801265	3.9e-8	0.16	Esophagus-Mucosa
		0.0000096	0.16	Skin (Not sun exposed)
Chr1:97450068A>G	rs17376848	No significant eQTL found		
DPYS				
Chr8:104466705G>A	rs2298840	No significant eQTL found		
Chr8: rs36027551G>A	rs36027551			
UPB1				
chr22:24495387A>T	rs2070475	2.10e-9	0.30	Adipose - Subcutaneous
		4.50e-8	0.30	Nerve - Tibial
		6.50e-7	0.57	Brain - Cortex
		6.90e-7	0.31	Thyroid
		2.4e-6	0.21	Whole Blood
		9.8e-6	0.44	Brain - Nucleus accumbens (basal ganglia)
		0.000012	0.31	Artery - Aorta
		0.000024	0.46	Brain - Putamen (basal ganglia)
chr22:24525761G>A	rs35916595	No significant eQTL found		
TYMS				
chr18:662247A>G	rs3786362	No significant eQTL found		

NES: Normalised Effect Size. A negative value indecates Downregulation and Positive value indecates Upregulation of the gene.

A



B

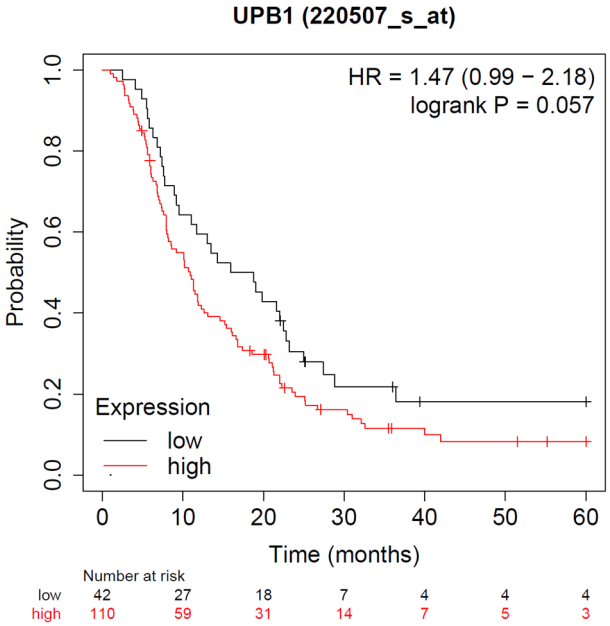


Figure 4.10: Survival analysis in KM Plotter. **A.** *DPYD* upregulation is associated with better overall survival (P=0.00047; HR=0.54), **B.** *UPB1* upregulation is associated with poor survival .

4.4.9 TCGA Data Analysis

DPYD, DPYS, UPB1, and *TYM genes* mutation frequencies were also iinvestigated in TCGA data (777 cases) hosted in cBioPortal. Among the four genes, *DPYD* was found to be mutated to a larger extent (4%) compared to *DPYS* (2.2%), *UPB1*(0.8%) and *TYMS* (0.1%) (Figure 4.11).

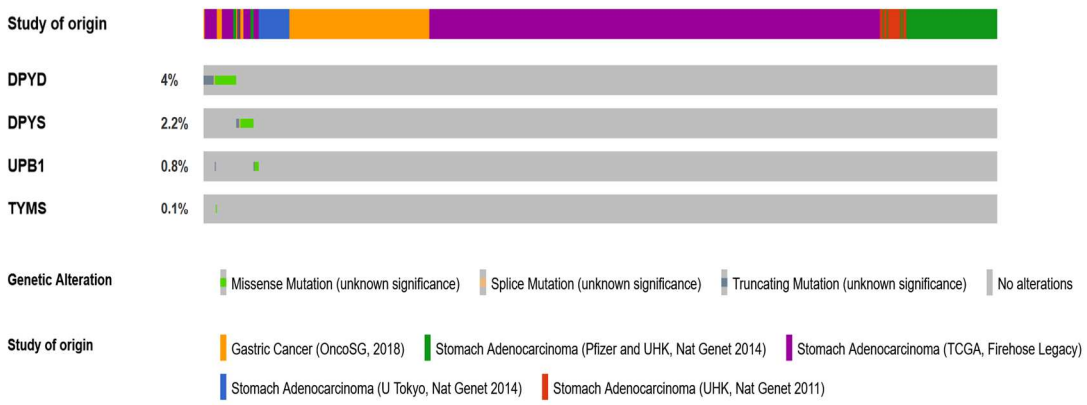
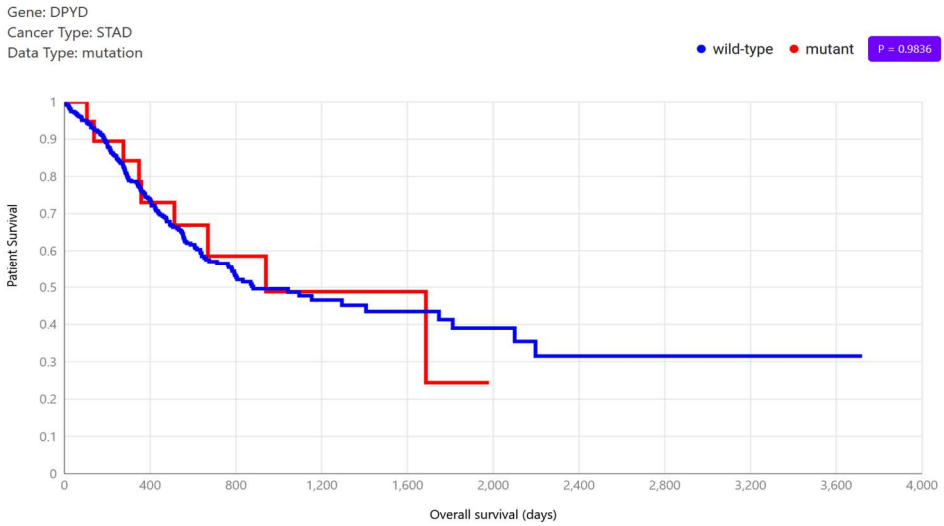


Figure 4.11: *DPYD, DPYS, UPB1* and *TYMS* gene mutation status in Stomach Cancer data hosted in cBioPortal.

4.4.10 Survival Analysis in TCGA Data

Kapler-Meier plots were analyzed in TCGA stomach adenocarcinoma (STAD) patients with mutation and corresponding survival data (https://www.tcga-survival.com). Patients with *DPYD* and *DPYS* mutations were found to have low survival rate. The patient with *DPYD* mutations survived less than 2000 days and *DPYS* mutations survived less than 2400 days) compared to the wild-type. However, these findings in the TCGA data were insignificant as the p-value for the survival of *DPYD* and *DPYS* mutants were more than 0.05 (Figure 4.12 A & B).

A



B

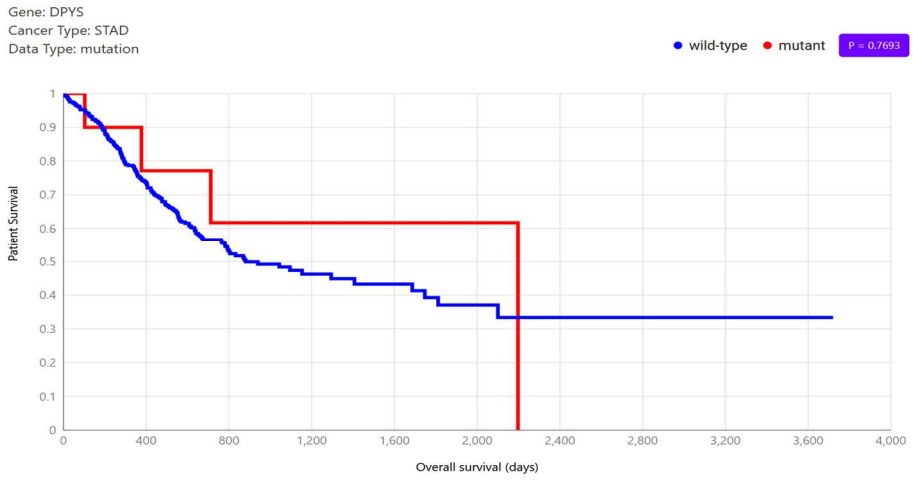


Figure 4.12: Kapler-Meier plots were analysed in TCGA stomach adenocarcinoma (STAD) patients with mutation and corresponding survival data: A. Survival of patient with *DPYD* mutation, B. Survival of patient with *DPYS* mutation

4.5 FLO toxicity-associated genetic variants

Three variants rs13181 in *ERCC2* (TG genotype), rs17376848 in *DPYD* (AG genotype), and rs1801133 in *MTHFR* (AG genotype) associated with FLO toxicity were found in 30.5%, 10% and 11.86% of GC patients, respectively. However, less toxic genotypes of the variants rs25487 in *XRCC1* (CT genotype), rs1695 in *GSTP1* (AG genotype), rs1799794 in *XRCC3* (CT genotype), rs717620 in *ABCC2* (CT genotype) and rs11615 in *ERCC1* (AG genotype) were present in 88.13%, 27.11%, 1.69%, 38.98% and 98.30 % of patients, respectively. The variant rs1799794 in *XRCC3* was not reported in ClinVar and there its pathogenicity and functional evidences are yet to be revealed. The rs13181 was reported as variant of unknown significance (VUS) and benign or likely benign in two different studies. The variants, rs1695 and rs717620 are reported as benign in ClinVar database. The variants reviewed by expert panel in ClinVar has extensive evidence of asociation with pathogenicity in term of toxicity and drug response in different other studies and are categorised as three star variants (Table 4.12 A & B).

However, these variants may exist alone or as group in the patients. It was observed that maximum variants occurred as a group in the patient. The variants rs13181 in 2 patients and rs11615 in 3 patients occurred as singleton. The variants rs11615, rs25487 and rs717620 occurred as together in 16 patients (27%). Moreover, rs11615 and rs25487 occurred together in 13 patients (22%) (Figure 4.11 A).

The prevelance of the variants rs13181 (30.50%), rs25487 (88.13%), rs1695(27.11%), rs717620 (38.98%) and rs11615 (98.30%) were high among the 59 gastric cancer patients. However, all the five variants was not seen together in any patient (Figure 4.11 B). However, the variants rs1801133, rs25487 and rs11615 which are reviewed by the expert panel in ClinVar have occurred in 6 Patients (10%). (Figure 4.11 C)

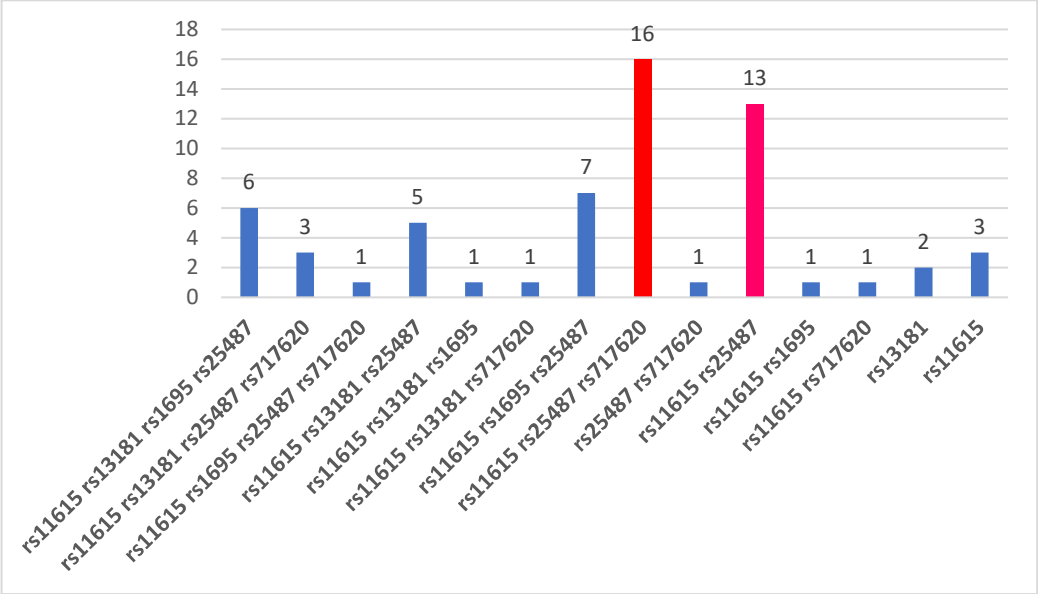
Table 4.12 A: FLO Toxicity Associated genetic variants in all the gastric cancer patients (*n=59*):

Variant ID	Toxic Genotype	Present Genotype	Gene	Variant Details	No. of Samples	Positive Samples (%) for the variants	ClinVar Status
rs13181	**TG**	**TG**	***ERCC2***	**NM_000400:exon23:c.A2251C:p.K751Q**	**18**	**30.50**	**VUS, B/LB**
rs17376848	**AG**	**AG**	***DPYD***	**NM_000110:exon14:c.T1896C:p.F632F**	**6**	**10**	**B**
rs1801133	**AG**	**AG**	***MTHFR***	**NM_001330358:exon5:c.C788T:p.A263V**	**7**	**11.86**	**Reviewed by expert panel**
rs1801131	GG	GT	*MTHFR*	NM_001330358:exon8:c.A1409C:p.E470A	16	27.11	**CIP**
rs25487	**CC**	**CT**	***XRCC1***	**NM_006297:exon10:c.A1196G:p.Q399R**	**52**	**88.13**	**Reviewed by expert panel**
rs1695	**GG**	**AG**	***GSTP1***	**NM_000852:exon5:c.A313G:p.I105V**	**16**	**27.11**	**B**
rs1799794	**CC**	**CT**	***XRCC3***	**NM_005432:c.-1843A>G (UTR5)**	**1**	**1.69**	**Not reported**
rs717620	**CC**	**CT**	***ABCC2***	**NM_000392:c.-24C>T**	**23**	**38.98**	**B**
rs11615	**AA**	**AG**	***ERCC1***	**NM_001369411:exon3:c.T354C:p.N118N**	**58**	**98.30**	**Reviewed by expert panel**

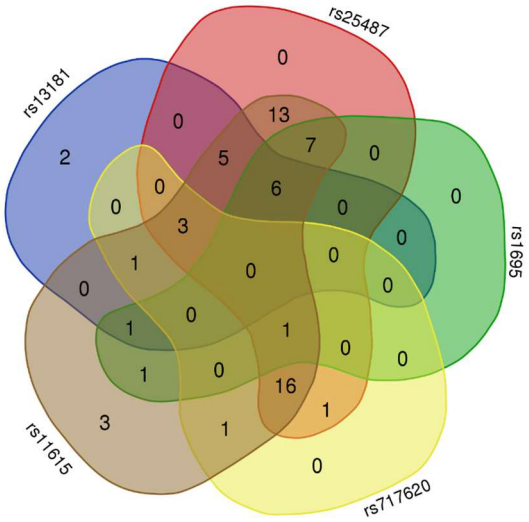
Table 4.12 B: FLO Toxicity Associated genetic variants with the sample IDs (*n=59*):

Variant ID	Sample IDs	No. of Samples	Positive Samples (%) for the variants
rs13181	gc-128, gc-139, gc-143, gc-14, gc-21, gc-41, gc-42, gc-47, gc-51, gc-54, gc-58, gc-50, gc-73, gc-75, gc-82, gc-83, gc-98, gc-9.	18	30.50
rs17376848	gc-133, gc-47, gc-64, gc-73, gc-85, gc-86	6	10
rs1801133	gc-41, gc-42, gc-58, gc-60, gc-70, gc-88	**7**	11.86
rs1801131	gc-126,gc-128,gc-136,gc-142,gc-143,gc-145,gc-148,gc-41,gc-47,gc-59,gc-60,gc-68,gc-75,gc-84,gc-86,gc-8	16	27.11
rs25487	gc-106,gc-126,gc-132,gc-133,gc-135,gc-136,gc-139,gc-13,gc-140,gc-141,gc-142,gc-143,gc-145,gc-147,gc-148,gc-17,gc-18,gc-29,gc-2,gc-35,gc-40,gc-41,gc-42,gc-47,gc-49,gc-4,gc-51,gc-54,gc-56,gc-58,gc-59,gc-5,gc-60,gc-64,gc-65,gc-67,gc-68,gc-70,gc-71,gc-73,gc-75,gc-79,gc-81,gc-82,gc-83,gc-84,gc-85,gc-86,gc-88,gc-8,gc-98,gc-9	52	88.13
rs1695	gc-135,gc-139,gc-13,gc-142,gc-147,gc-148,gc-21,gc-2,gc-47,gc-5,gc-63,gc-67,gc-73,gc-75,gc-83,gc-98	16	27.11
rs1799794	gc-13	1	1.69
rs717620	gc-106,gc-128,gc-131,gc-132,gc-135,gc-140,gc-141,gc-17,gc-18,gc-35,gc-40,gc-41,gc-49,gc-4,gc-54,gc-59,gc-64,gc-65,gc-79,gc-81,gc-82,gc-88,gc-8	23	38.98
rs11615	gc-106,gc-126,gc-128,gc-130,gc-131,gc-132,gc-133,gc-135,gc-136,gc-139,gc-13,gc-140,gc-141,gc-142,gc-143,gc-145,gc-147,gc-148gc-17,gc-18,gc-21,gc-22,gc-29,gc-2,gc-35,gc-40,gc-41,gc-42,gc-47,gc-49,gc-4,gc-51,gc-54,gc-56,gc-58,gc-59,gc-5,gc-60,gc-63,gc-65,gc-67,gc-68,gc-70,gc-71,gc-73,gc-75,gc-76,gc-79,gc-81,gc-82,gc-83,gc-84,gc-85,gc-86,gc-88,gc-8,gc-98,gc-9	58	98.30

A



B



C

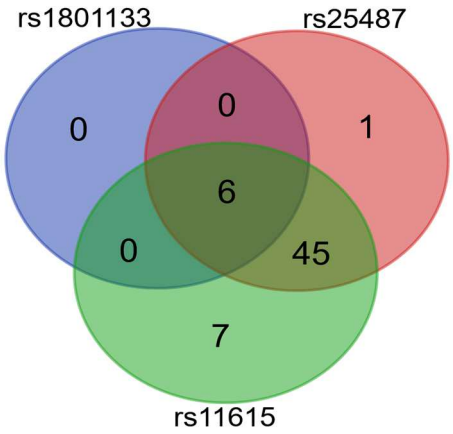


Figure 4.11 : Co-occuernce of the FLO toxicity associated variants: **A.** Number of patient have the variants as groups or the variant alone, **B.** Co-occuerance of higly prevalant variants where no indivisual got the five variants together as a group, **C.** The variants reviewd by the expert panel (Three Star Variants) co-occurred in the six patients (gc-60, gc-88, gc-41, gc-70, gc-42, and gc-58).

CHAPTER 5
DISCUSSION AND CONCLUSION

5.1 PGx-Gene Variants in Healthy Individuals from Northeast India

The PharmGKB database has classified 67 genes as very important pharmacogenes (VIPs). These VIPs were further categorized into Tier 1 with 33 genes, Tier 2 with 25 genes, and Tier 3 with 9 genes. The number of VIP genes has recently increased to 68 genes with the addition of *TYMS* and is expected to grow more based on the role of the genes in clinically important therapeutics (Hewett et al., 2002).

The 67 genes were screened for the variants in healthy exome and NE Indigenomes datasets. In the exome datasets derived from 27 healthy volunteers, there were 490 variants within the coding, 1112 in intronic, 95 in 3'UTR, and 52 in 5'UTR regions. Among the variants, 163 variants were non-synonymous, and 315 variants were found to be synonymous. Moreover, 1 frame-shift substitution, 4 start-loss, 1 stop-loss, and 2 stop-gain (2) variants were found. About 77,799 variants were found in 67 VIPs in the NE IndiGenomes dataset. There were 2527 variants in the exonic region, whereas 67,144 variants were found in intronic regions. There were 2086 variants in 3'UTR3, 378 variants in 5'UTR, 107 variants in intergenic, 82 variants in ncRNA_exonic, 5445 variants in ncRNA_intronic, 24 variants in splicing, 5 variants in upstream, and 1 variant in exonic splicing region. Among the other variant types, 910 variants were synonymous, and 1462 variants were non-synonymous. Moreover, 59 frameshift, 40 non-frameshift, 8 start-loss, 3 stop-gain, stop-loss variants were also found. However, 8 variant types remained unknown. The visualization of the variant details is made available through **M**izoram **P**harmacogenomics **V**ariant **D**atabase MPVardb v1.1 constructed using MariaDB v10.4.28 with XAMPP v 3.3.0. The database query in the form of the HGNC gene symbol (among 67 VIPs) fetches the data from the database and displays the results using PHP. Visualization of the distribution of variants types from the healthy exome and NE Indigenomes datasets are achieved with the help of interactive pi-charts that were made available through Java script with the PHP result page. The home page of the database was designed using HTML and CSS.

A vast difference in the number of variants in exonic variants in both datasets was observed. The exonic variants in exome and NE Indigenomes datasets were found to be 490 and 2527 variants, respectively. Although exome represents only 2% of the genome and WES technology can sequence those coding regions, one of the major limitations of WES is the uneven coverage of the target regions by the sequence reads (Wang et al., 2017). This could be a major reason for a large number of exonic variants in the NE IndiGenomes datasets compared to the healthy exome, however, the uneven sample size and population diversity cannot be ignored. Currently, MPVardb hosts all types of germline variants including novel variants and rare variants (MAF < 0.01) resulting from the WES and WGS experiments based on the Northeast Indian populations. The database provides the allele frequencies of the variants derived from the healthy exome from Mizoram (n=27) and from the NE Indigen datasets (n=93). The allele frequency derived from the exome data sometimes may not explain properly the true prevalence of the variants within the population or in the region. Allele frequencies obtained from WES and WGS can be diverse due to inherent differences in their methodologies. WES selectively captures and sequences only coding regions, potentially missing variants in non-coding areas that may influence allele frequencies. WGS, on the other hand, encompasses the entire genome, offering broader coverage. Technical factors such as sequencing errors, biases, and variant calling algorithms can also introduce disparities. Additionally, rare variants and population-specific differences can play a role, as can sampling variability when dealing with small sample sizes.

Genomic databases are one of the major parts of genomic research as it allows to get access to genetic variants from different population or ethnicity. There are a number of such databases that preserve the genetic variants as well as their clinical importance. However, the variants and their impact on health vary from population to population. It is extremely important to study genetically isolated populations to understand their impact on health (Król et al., 2023). This is the first work done for cataloging the variants from northeastern populations, primarily the Mizo population as the region is known to be genetically diverse. The database, aside from its data storing and visualization functions also holds the potential to enhance diagnostics and

treatment methods. The number of variants in **MPVardb** is expected to grow in the future upon recruiting more volunteers from the region to understand the pharmacogenomic landscape to personalize therapies.

5.2 Clinically Actionable Variants of PGx-Genes in Healthy Exomes

The genetic variations in genes involved in ADME alter the drug effect and provoke ADRs in patients. Pharmacogenomics (PGx) plays a crucial role in precision medicine in order to minimize ADRs and provide effective therapy. Clinical pharmacogenomics can be utilized to understand the allelic variation to tailor personalized therapy based on the individual genomic profile. However, the task is extremely challenging due to the lack of information on the allelic variation that affects the drug action. A large-scale genomic study helps identify the prevalence the PGx-gene variants and their probable impact based on international guidelines might be one of the promising steps to predict the drug effect within the population and also to design an effective therapy (Daneshi et al*.,* 2023). One of the noteworthy advancements in this field has been the establishment of comprehensive clinical guidelines for drug-gene interactions. These guidelines have primarily evolved through collaborative efforts such as the Clinical Pharmacogenetics Implementation Consortium (CPIC) in the United States, European counterpart known as the Dutch Pharmacogenetics Working Group (DPWG), and the invaluable PharmGKB PGx database (Nunez-Torres et al., 2023).

Through this study, it was aimed to identify and categorize specific genetic variations with significant clinical relevance. The utilization of star nomenclature allowed for a precise and systematic approach to characterizing these variants, ultimately contributing valuable insights into their potential impact on medical decision-making and patient care.

Many of the actionable alleles consist of multiple variants, posing a complex challenge when analysing variants from genotype/sequencing data from both chromosomes concurrently. The resolution of this intricate issue is necessary on a gene-by-gene basis as per the CPIC guidelines (Klein et al*.,* 2018). Frequently, CPIC is often contacted by users of guidelines to assemble a list of variants that hold clinical

significance. These clinically relevant variants refer to those that when found in the appropriate gene combination (often as part of a diplotype along with another similarly actionable variant) have the potential to influence prescribing decisions away from the standard course of action (Relling et al., 2018).

rs2231142 (*ABCG2*)

ABCG2 is responsible for developing resistance against cancer therapy. There are a number of chemotherapeutic agents such as imatinib, doxorubicin, and mitoxantrone are substrates to *ABCG2* (Stacy et al., 2013). It was found in a study that genetic variant rs2231142 of *ABCG2* also reduces the effectiveness of allopurinol, a drug used to lower uric acid levels in the blood, necessitating higher dosages. This variant is responsible for decreasing in uric acid excretion by the kidneys and intestines which causes a rise in uric acid concentration in blood. A higher than standard dose of allopurinol is necessary to achieve the required inhibition and reduce the effect of the variant. Although CPIC has not provided any guidelines yet, DPWG recommends using 1.25 times the standard dose of allopurinol (Wen et al., 2015). The allele frequency (AF) in the healthy exome and Indigen (NE datasets) projects were found to be 0.259 and 0.811, respectively.

CYP2B6*2

CYP2B6 is an important gene responsible for drug metabolism belonging to P450 in the liver. The gene metabolizes a number of drugs that include artemisinin used for the treatment of malaria, bupropion used as an anti-depressant, cyclophosphamide used as a chemotherapy against breast cancer, efavirenz used against HIV, ketamine used to maintain anesthesia, and methadone used to treat opioid addiction. *CYP2B6* is considered as one of the most highly polymorphic genes for CYP family in human (Zanger et al., 2013). However, the CPIC and DPWG have not provided any guidelines for the particular variant *CYP2B6*2* (rs8192709). Moreover, there are no evidence on *2 allele's effect on drug metabolism. The variant AF in healthy exome and IndiGen NE datasets were found to be 0.111 and 0.913, respectively.

TPMT*3A

TPMT is another important pharmacogene that encodes for thiopurine methyltransferases. The enzyme metabolizes thiopurin-based chemotherapeutic drugs. The variant Chr6:18130687 T>C (*3A) causes deficiency in the enzyme slowing down the metabolism of thiopurine-based drugs and subsequently increasing the levels of TGN metabolites and low levels of MeTIMP (Hindorf et al., 2006; Skrzypczak-Zielinska et al., 2013). It has been reported that deficiency in the enzyme also causes thiopurine-associated leukopenia characterized by lower white blood cells, neutropenia characterized by low levels of neutrophils, and myelosuppression characterized by lowering the ability of bone marrow to make platelets. CPIC and DPWG have both provided guidelines for the *TPMT*3A* variants for dose adjustment of thiopurine based drugs. As per the CPIC guidelines, starting therapy with reduced initial doses, typically ranging from 30 to 80% of the standard starting dose, which is typically set at approximately 2-3 mg/kg/day. Azathioprine dose modifications should be guided by the extent of myelosuppression and adhere to disease-specific protocols. It is crucial to allow a 2 to 4-week interval for the medication to attain a stable state after each adjustment (Dean et al., 2012). The DPWG says that 23% of patients might get a condition called leukopenia (which is when your white blood cell count is too low) if they take certain medications while they have a normal immune system. Some people have genes that make these medications work more strongly in their body and it is good for those people to start with only half of the usual dose of the medicine. The doctors should monitor their blood counts and how well the medicine is working to make sure it's safe and effective. But if the dose is lower than 1.5 mg/kg per day for one of the medicines (azathioprine) or 0.75 mg/kg per day for the other one (mercaptopurine), one doesn't need to adjust it (Swen et al., 2011). The variant AF was found in the healthy exome was 0.01 and NE Indigen data was 0.961.

***G6PD* (Mahidol)**

People who have a certain genetic variant have a little risk of getting a type of anemia called acute hemolytic anemia. The *G6PD* Mahidol variant also exhibited an approximately 40% reduction in enzyme activity compared to individuals with the

wild-type version. However, it was also reported that the mahidol variant also carry a protective effect against *Plasmodium vivax* Kachin population from the northeastern part of Myanmar (Yi et al., 2019). *G6PD* deficiency renders red blood cells exceptionally prone to oxidative harm, making them more susceptible to a process known as hemolysis, where they break down (Luzzatto et al., 2020). Despite reduction of enzyme activity, CPIC suggests that there's no need to avoid certain medications because of this genetic difference when taking the usual doses. However, for people with more than one X chromosome (Female) who have one normal version of the gene and one with the variant allele, their condition might look normal or deficient because of a mix of these genes (Gammal et al., 2023). The AF in healthy exome and NE IndiGen dataset was found to be 0.05 and 0.989, respectively.

UGT1A1*6

This genetic difference causes a slight decrease in the activity of the *UGT1A1* enzyme, but it's unlikely to make patients stop taking atazanavir because of bilirubin-related side effects. Therefore, it is not necessary to stop prescribing atazanavir by clinicians based on the status of the variant. It should not be overlooked that some people have to stop taking atazanavir because it can make their skin and eyes turn yellow (Gammal et al., 2016). Most of the studies linking *UGT1A1* genes with atazanavir side effects have looked at the use of ritonavir to boost the medication. However, when cobicistat is used for boosting instead of ritonavir, the levels of the medication in the body are similar (Gallant et al., 2013). This suggests that bilirubin-related side effects, including the need to stop atazanavir, are likely to be similar when atazanavir is taken with cobicistat as well. The variant AF in healthy exome was found to be 0.13 and in NE indigen datasets was found to be 0.779.

SLCO1B1*14

The optimum way to start with the dose that's typically recommended and then alter as needed according to guidelines specific to the patient's condition. Before starting on statin therapy (a type of medication), it's crucial to check if there could be any interactions with other prescribed drugs for the same patient. Also, the kidney and liver functions as well as the ancestry have to be considered when deciding on the

maximum safe dosage. The variants AF in healthy exome and IndiGen NE datasets were found to be 0.722 and 0.354, respectively.

DPYD*5

DPYD is one of the important genes that play a key role in 5-Fluorouracil metabolism. There are a number of variants already documented that are known to cause DPD deficiency. *DPYD*5* was observed in both healthy exome and NE indiGen datasets with allele frequencies 0.225 and 0.725, respectively. CPIC and DPWG both have not issued any guidelines for the variant on treatment of cancer with 5FU. However, it has been found that *DPYD*5* causes reduced activity of DPD enzyme and induces toxicity (Zhang et al., 2007; The et al., 2013). Previous studies on *DPYD*5* have revealed that *DPYD*5* showed a weak association with activity of the DPD enzyme in African Americans (Offer et al., 2013).

rs17376848 (*DPYD*)

The rs17376848 is another common variant of *DPYD*. The variant allele frequency was found in healthy exome and NE IndiGen was found to be 0.111 and 0.924, respectively. This variant can increase the risk of severe side effects when individuals with this genetic variation are treated with these drugs. In a study, the variant was found associated with grade 3 toxicity of 5FU within first three cycles (Toffoli et al., 2015; Puerta-García et al., 2022). The variant was also found to be a predictive biomarker for 5FU response in cancer patients (The et al., 2013).

Currently, the variants *DPYD*5* and rs17376848 (*DPYD*) have not yet received clinical classification as actionable variants, and neither the CPIC nor DPWG has issued guidelines recommending dosage adjustments for 5FU and its analogs based on these variants. Nonetheless, growing evidence of toxicity linked to *DPYD*5* is raising concerns that may warrant further investigation.

It was observed in the study that the allele frequencies of the key star alleles are less in Mizo healthy exomes compared to NE IndiGen datasets. The study uncovered differences in the distribution of star allele frequencies between Mizo healthy exomes and the NE IndiGen datasets. Such disparities in allele frequencies

across distinct populations are typically shaped by a complex interplay of genetic, historical, and demographic factors (Giuliani et al., 2018), necessitating an in-depth study in the Mizo population. Genetic diversity is a pivotal contributor to the observed variations in allele frequencies (Morales-González et al., 2021). Like many indigenous groups, the Mizo population may have evolved in relative isolation, giving rise to unique genetic profiles. Conversely, the NE IndiGen dataset is likely more diverse, comprising individuals from various ethnic backgrounds in Northeast India. Consequently, the genetic diversity within the Mizo population may be reduced, leading to lower frequencies of specific star alleles.

Founder effects could also be influencing the observed differences. Historically, the Mizo population may have experienced genetic bottlenecks or founder effects, scenarios where a small group of individuals establishes a new population. Such events often lead to reduced genetic diversity in specific alleles. In contrast, the NE IndiGen dataset with its broader representation may not exhibit such constraints on allele frequencies. Geographic isolation is another plausible factor contributing to the disparities in star allele frequencies (Gayden et al., 2009). The Mizo population may have limited gene flow with neighboring communities or regions, resulting in the development of distinct genetic profiles over time. In contrast, the NE IndiGen dataset represents a broader and more interconnected genetic landscape.

Selective pressures might account for the observed differences in allele frequencies. Local environmental factors, dietary practices, or disease prevalence may have favored or disadvantaged certain alleles within the Mizo population, leading to variations in the prevalence of specific star alleles compared to the NE IndiGen dataset. Sampling bias should not be overlooked when interpreting these findings. The composition of the Mizo healthy exomes dataset may not fully represent the entire Mizo population. In contrast, the NE IndiGen dataset's larger and more diverse representation could skew allele frequencies. Ensuring unbiased and comprehensive sampling is crucial for accurate genetic comparisons.

Historical migration patterns can significantly impact allele frequencies. The Mizo population's unique migration history or prolonged isolation may have led to

distinctive allele frequencies. Conversely, the NE IndiGen dataset could include populations with different migration histories, further contributing to the observed differences (Pachau et al., 2022). Small population size is another factor influencing the disparities in star allele frequencies. In smaller populations, genetic drift can play a substantial role in allele frequency fluctuations (Maruyama et al., 1985). Random events can lead to the loss or fixation of specific alleles, contributing to differences between the Mizo healthy exomes and NE IndiGen datasets.

Genetic drift and founder effects can further compound these disparities. Over time, in smaller populations, random fluctuations in allele frequencies can become more pronounced, potentially leading to the observed differences. Founder effects, which occur during the initial establishment of a population, can also amplify these disparities. To ascertain the precise factors driving these variations in star allele frequencies, additional genetic research, population genetics analyses, and historical investigations are warranted. Furthermore, conducting functional studies can help elucidate whether these differences have any biological significance in terms of health or adaptation within these distinct populations.

5.3 Adverse Drug Reaction in Gastric Cancer

Adverse Drug Reactions (ADRs) significantly contribute to illness and death, imposing a substantial economic burden on both individuals and society (Sharma et al., 2015). In this study, 37 gastric cancer patients were followed up and 29 ADRs of different grades were documented. The portion of males and females showing different ADRs were 75.68% and 24.32%, respectively. Loss of Appetite was found to be a priominent ADR which was observed in 26 patients, followed by the occurrence of Fatigue in 21 patients, Nausea in 23 patients, drowsiness in 18 patients, Dermatological Adverse Reaction in 11 patients, Vomiting in 8 patients, Alopecia in 8 patients, and many more. Neutropenia, as often can be seen as ADR in patients undergoing anti-cancer therapy was also observed in 3 patients. Similarly, Leukopenia in 2 patients and low hemoglobin count in 1 patient. However, these ADRs were not observed as singletone. The patient showing Loss of Appetite also showed other ADR as comorbidities such as nausea, and drowsiness with others.

Numerous studies have demonstrated that chemotherapy can lead to the occurrence of adverse drug reactions (ADRs) in patients, which can have a significant impact on their well-being. In gastric or stomach cancer, the most commonly reported ADRs are related to gastrointestinal problems such as nausea, vomiting, constipation, and loss of appetite. However, a proportion of patients also experience side effects like hair loss (alopecia) and hematological problems (Chopra et al., 2016; Wahlang et al., 2016). Chemotherapy can be effective in killing the remaining cancer cells after surgery or slowing down cancer progression, the ADRs due to therapy may vary in severity and frequency among individuals. Studies have found that genetics plays a important role in determining a patient's susceptibility to toxicity induced by chemotherapeutic drugs (Franczyk et al., 2022). Specifically, deficiencies in enzymes or differing in activity of the enzymes involved in the metabolic pathways of these anti-cancer drugs can lead to ADRs. Additionally, it's important to consider the potential influence of drug-drug interactions, as they can also contribute to the occurrence of ADRs (Sahana et al., 2021; Franczyk et al., 2022). Understanding these factors is crucial for healthcare professionals in optimizing chemotherapy regimens, minimizing ADRs, and improving the overall quality of care for cancer patients.

It is essential for healthcare providers to closely monitor patients, manage side effects proactively, and adjust treatment when necessary to optimize the therapeutic benefits while minimizing discomfort and risks associated with these reactions. Effective communication between patients and their healthcare team is vital in addressing and managing these challenges, ensuring the best possible outcomes in the fight against stomach cancer. Furthermore, pharmacogenomics research can identify the potential genetic variants that induce such toxicities in patients.

There is a need to closely examine how side effects of medicines are assessed during clinical trials. Although there's significant attention given to monitoring side effects after a drug is on the market, a well-structured system for evaluating them during earlier testing stages is lacking. Such a system would help in better detecting and measuring side effects consistently and reliably from the beginning of testing until the drug becomes available to the public. This would ensure greater safety for patients using these drugs, provide doctors with more information about the side effects of

medicines that affect the mind, and enhance the ability to compare experimental results conducted at different testing centers.

Given the substantial impact of severe ADRs, policymakers should consider the establishment of independently funded pharmacovigilance centers of excellence to aid clinician investigations.

5.4 Genetic Variants in PK/PD pathway of fluorouracil

The study found that most gastric cancer patients (about 92%) at least 5FU and 72% of it recieves 5FU with leucovorin and oxaliplatin. This made it really important to check for certain gene variations that affect how the body metabolizes 5FU, as 5FU undergoes enzymatic breakdown to exert its effect. On the other hand, oxaliplatin undergoes non-enzymatic breakdown. These gene variations can impact how effective the treatment is and whether there might be any side effects. Therefore, genetic screenings can provide valuable insights into individualized therapy approaches, ensuring that patients receive the most suitable and tailored treatments while mitigating potential adverse drug reactions, thereby advancing the field of gastric cancer management.

Three specific genes, *DPYD*, *DPYS*, and *UPB1*, play essential roles in the way the body processes and responds to 5-Fluorouracil (5FU). *DPYD*, for example, is responsible for breaking down 5FU, and variations in this gene can affect how quickly or slowly the drug is metabolized, potentially leading to variations in its effectiveness and the risk of side effects. *DPYS* and *UPB1* also contribute to the complex process of how 5FU interacts with the body's biological pathways. Moreover, the *TYMS* gene plays an important role in the pharmacodynamics of 5FU. Understanding these genetic factors is crucial in tailoring 5FU treatments to individual patients, ensuring optimal therapeutic outcomes while minimizing potential adverse reactions.

The current study found the involvement of these three genes in 5FU metabolism based on the literature. It has been cited that deficiency in any of these three genes may exert 5FU-related adverse reactions (https://www.pharmgkb.org/; Hewett et al., 2002). These three genes were screened in 59 patients' annotated data for synonymous and nonsynonymous variants associated with deficiency of the

respective enzymes. A total of 5 variants in *DPYD* and 2 variants in each in *DPYS* and *UPB1* among 59 patients. Mutation status in the publicly available data in cBioPortal

5.4.1 Variants in *DPYD* and their Impact

Exome sequencing revealed that 12 patients (52.17%) out of 23 patients who received 5-Fluorouracil had *DPYD* variants responsible for DPD Deficiency. Among the *DPYD* variants, rs1801159 (*DPYD*5*), a nonsynonymous variant was found in 35% of the total number of patients. Among the patients who received fluorouracil with exome sequencing done (*n=23*), 7 patients were positive for rs1801159. Five of these patients (accounting for ~ 21.74% of the patients) were deceased and 2 patients were found alive. The variant is reported in ClinVar as being responsible for DPD deficiency and toxicity associated with 5-Fluorouracil . Four out of the five deceased patients received 5-Fluorouracil with leucovorin and Oxaliplatin (6 to 8 Cycles) and one patient received oral capecitabine (4 Cycles).

The follow-up study was able to receive survival data from 3 deceased patients. One patient survived for less than a year, second patient survived less than two years and the third patient survived for less than three years. Six of these patients received intravenous 5FU with leucovorin and oxaliplatin and one patient received oral Capicitaben. The patient who received capecitabine showed grade 3 constipation. Loss of appetite of varying degrees was observed in 5 patients. Fatigue of grade 2 was noted in 3 patients, grade 1 in one patient, and grade 2 fatigue in 3 patients. Additionally, two patients displayed grade 2 alopecia. Grade 2 nausea was documented in two patients, while one patient experienced grade 3 nausea, and another patient had grade 1 nausea. Additionally, one patient displayed mild diabetes, and another patient exhibited grade 3 thrombocytopenia, which is characterized by low blood platelet levels. Constipation was observed in only one patient. Mild to moderate drowsiness was observed in 3 patients. Similarly, mild to moderate vomiting was observed in two patients, and severe vomiting in one patient. No significant cis-eQTL was found for the variant rs1801159 which indicates that the variant may not have impact on the expression of the gene.

Although PharmGKB has depicted the variant of normal function, our study found the variant has a great impact on patients' health. Moreover, a study conducted by Zhang et al. (2007), polymorphisms of *DPYD*5* (rs1801159) was over represented in non-responsive of fluorouracil treated patient from Chinese population and suggested that *DPYD*5* as probable predictors of the response to fluorouracil-based chemotherapy for gastric cancer patients. Another study found that 29.9% of the neutropenia cases were *DPYD*5* positive (ANOVA, P-Value = 0.01). The study also suggested that *DPYD*5* (rs1801159) is a potentially useful predictive markers of patients' responses to 5FU chemotherapy (Teh et al., 2013). Therefore, a statistical test will be necessary in this particular population for their association with the adverse drug reaction.

A rare nonsynonymous *DPYD* variant rs112766203 was detected in one patient, which was also reported to be associated with DPD deficiency in ClinVar. In the patient, there were various ADRs of different grades namely: Loss of Appetite (Grade 1), Nausea (Grade 1), Darkening of Nails (Grade 2), severe neutropenia, and Alopecia (Grade 1). The patient survived 512 days (17 months and 10 days) from the date of detection of cancer. The variant was predicted to be damaging by SIFT, Polyphen, and Mutation taster. The variant was also designated as probably damaging by Marieke et al. (2019). However, there is not enough evidence on the association of adverse events due to the variant in published literature.

Similarly, another nonsynonymous variant of *DPYD* was rs1801160 (*DPYD*6*) observed in one patient. In a study conducted by Matáková et al. (2017), the variant was found to be associated with colorectal cancer. The variant was also found to be responsible for slowing down the degradation rate of 5FU (Gentile et al., 2016). Another study revealed that the variant significantly induces fluorouracil-associated hematological toxicity in European patients (Kim et al., 2022). A recent study by Božina et al. (2022) suggested that the variant can be a potential candidate for the *DPYD* testing panel due its association with severe adverse drug reactions due to 5FU treatment. The study also found two synonymous variants of *DPYD*, rs1801265, and rs17376848 in two patients. Although, the patients were found alive, both the variants were reported to be associated with DPD deficiency and the variants

were also found to be associated with inducing ADRs in 5FU-treated patients in other studies (Teh et al., 2013; Ruzzo et al., 2017; Puerta-García et al., 2020; Hamazic et al., 2021).

The eQTL analysis revealed that variants rs1801160 and rs1801265 also cause upregulation of *DPYD* and the patients who showed positive for the variants were also found alive. Gene expression-based survival analysis using GEO datasets in KM-Plotter revealed that upregulation of *DPYD* was associated with better overall survival of the patient (p-value=0.00047, HR=0.54). However, due the less allele frequency of these variants (less than 0.05), the variants may not be suitable for *DPYD* testing for the population.

Based on the mutation-based survival study based on TCGA data revealed poor survival outcomes with the patient with the *DPYD* mutant allele. Moreover, looking at the AF of the variant rs1801159 in the 1000g, Mizoram healthy, NE IndiGenome as well as the occurrence of the variant in gastric cancer patients, the study suggests in-depth clinical research to generate evidence of association with ADRs and DPD enzyme activity which could serve as potential predictor for the response of 5FU treatment that are being provided to the stomach cancer patients in the state of Mizoram.

5.4.2 Variants in *DPYS* and their Impact

Two synonymous variant of *DPYS* gene found in the followed up patients. The variant rs2298840 was detected in 11 out of 23 patients and the variant rs36087551 was detected in one patients along with rs2298840. Among the 11 patients, 7 patients (63%) weere found deceased. The variant rs2298840 was also found in 46% of patients out of all the 59 GC patients. The patient with the variant rs36027551 was also deceased and variant was not observed in the 1000g annotation datasets. The AF of the variants rs2298840 in 1000g datasets was 0.231629. The AF of the variants rs2298840 in 1000g datasets was 0.231629 (Table B). One patient (gc-35) had two *DPYS* variants rs2298840 and NM_001385:exon3:c.C541T:p.R181W (rs36027551) along with one *DPYD* variant rs1801159 that are associated with DPD deficiency and Dihydropyrimidinase deficiency found to be diseased. It was observed that the

majority of the deceased patients (n=6) survived less than 5 years (1825 days). Although some of the variants of *DPYS* were found associated with the 5FU toxicity, there was no evidence of association the variants rs2298840 and rs36027551 with toxicity. Moreover, no significant eQTL was found for these variants. *DPYS* mutation status was found to be 2.2% in the stomach cancer data available in cBioPortal. The variant rs2298840 was found in 11 patients of which 7 patients (~63%) did not survive. This finding led us to check the role of *DPYS* mutation in survival in TCGA data. It was observed that *DPYS* mutation might play role in poor outcome in gastric cancer patients. These findings necessiates further in-depth study on *DPYS* variants in stomach cancer patients treated with 5FU.

5.4.3 Variants in *UPB1* and their Impact

Only one exonic variant (synonymous) NM_016327:exon10:c.G1122A:p.K374K (rs35916595) was observed in the *UPB1* in the heterozygous condition in one patient (gc-79). The variant is reported for deficiency in bete-ureidopropionase, an enzyme that converts luoro-beta-ureidopropionate to fluoro-beta-alanine (FBAL). The AF of the variant in 1000g datasets was found to be 0.0042. Another variant in 5'-UTR NM_016327:c.-17A>T (rs2070475) was also in the same patient. The patient experienced grade 3 mucositis with neutropenia and survived for 548 days from the date of detection of the disease. Moreover, two patients (gc17 and gc-58) carrying the variant rs2070475 survived for 1665 days and 875 days, respectively. Different types of adverse reactions were observed in these patients including severe forms of anorexia, SOB, Diarrhoea and vomiting in gc17, mild forms of LoA, Anorexia, and diabetes in gc-58 and low haemoglobin (grade 2), mucositis (grade 3), and neutropenia (grade 3) in gc-79. Overall, 3 patients were found to be deceased out of 6 patients with *UPB1* variants and 3 patients are still surviving. One of the alive patients (gc-83) experienced LoA (grade 3), fatigue (grade 2), nausea (grade 3), drowsiness (grade 3), DAR (grade 3), Alopecia (grade 3). The other alive patient experienced LoA (grade 2), Diarrhoea (grade 2), DAR (grade 3) and Alopecia (grade 2) ADRs. Interestingly, significant eQTL were found to be associated with the variant rs2070475. The variant caused upregulation of the *UPB1* in many tissue types. To understand role of upregulation of *UPB1* on

survival, GEO datasets in KM-plotter were explored which suggested that the upregulation is not associtated with overall survival of the patients (P=0.057, HR=1.47). Furthermore, the survival data was not found in TCGA based on *UPB1* mutation status. The mutation frequency of the gene is also less (0.8%) in the stomach cancer data available.

5.4.4 ADRs in patients with more than one PK/PD gene variants

Furthermore, it was found that various ADRs were seen in the patient having more than one PK/PD gene variants. Grade 3 osteoporosis as ADR was observed in one patient (gc-2) and Loss of Appetite (grade 2), Fatigue (grade 2), and Nausea (grade 2) in another patient with variants rs1801159 in *DPYD*, rs2298840 in *DPYS*, and rs3786362 in *TYMS*. Loss of Appetite (grade 2), Vomiting (grade 1), Dermatological Adverse Reaction (Grade 1), and Alopecia (grade 2) were observed in a patient (gc-35) with two synonymous variants in *DPYS* (rs36027551 and rs2298840) and nonsynonymous variants in DPYD (rs1801159). Constipation was observed in one patient with rs1801159 in *DPYD* and rs2298840 in *DPYS.* Similarly, two patients (gc 75 and gc76), rs3786362 in *TYMS* and rs1801159 in *DPYD* had experienced more than three ADRs. In one patient (gc-75), the following symptoms were noted: grade 2 Vomiting, Fatigue, Drowsiness followed by grade 3 LoA, Nausea, DAR, Hypertension and grade 1 diabetes. Another patient (gc-76) exhibited milder symptoms, including grade 1 LoA, Fatigue, Nausea, and Drowsiness.

5.5 FLO toxicity-associated variants

Three variants rs13181 in *ERCC2* (TG genotype), rs17376848 in *DPYD* (AG genotype), and rs1801133 in *MTHFR* (AG genotype) associated with FLO toxicity were found in 30.5%, 10% and 11.86% of GC patients, respectively. However, less toxic genotypes of the variants rs25487 in *XRCC1* (CT genotype), rs1695 in GSTP1 (AG genotype), rs1799794 in *XRCC3* (CT genotype), rs717620 in ABCC2 (CT genotype) and rs11615 in ERCC1 (AG genotype) were present in 88.13%, 27.11%, 1.69%, 38.98% and 98.30% of patients, respectively. The variants rs13181 in 2 patients and rs11615 in 3 patients occurred as singleton. The variants rs11615, rs25487 and rs717620 occurred as together in 16 patients (27%). Moreover, rs11615 and

rs25487 occurred together in 13 patients (22%). The prevelance of the variants rs13181 (30.50%), rs25487 (88.13%), rs1695(27.11%), rs717620 (38.98%) and rs11615 (98.30%) were high among the 59 gastric cancer patients. However, all the five variants was not seen together in any patient. However, the variants rs1801133, rs25487 and rs11615 (reviewed in ClinVar) have occurred in 6 Patients (10%). Since these variants were already found to be associated with FLO toxicity in published literature and were also present in a group in the stomach cancer patient, further study is required to understand their role in developing toxicity.

5.6 Limitations of the Study

It's important to recognize that these studies have a limitation: there are not many clinical experts who are experienced in using pharmacogenetics. Moreover, due to the absence of electronic data capture medium for adverse events, it was challenging to collect all kinds of real-time adverse events during the therapy. A substantial number of patients were untraceable which resulted in a smaller number of samples and due to which a statistical analysis could not be performed to identify potential genetic factors for ADR.

CHAPTER 6
INCIDENTAL FINDING

6.1 Case representation

In 2016, a 58-year-old woman from Mizoram state of India, had been diagnosed with ER^+ metastatic breast cancer (well-differentiated, ductal carcinoma, stage: III A and TNM: $T_1N_3M_0$), underwent a modified radical mastectomy followed by a 4 cycles of chemotherapy regimen consisting of 5-Fluorouracil, Epirubicin, and Cyclophosphamide. Following the completion of her treatment, the patient exhibited several notable medical findings. These included the identification of a lump in the right axilla, a cystic nodule located in the left thyroid lobe, liver cysts, and the a pulmonary nodule. These diverse manifestations indicate the emergence of abnormal growths or masses in distinct anatomical regions, necessitating further investigation and medical attention to ascertain the underlying causes and implement appropriate therapeutic measures.

Further complicating her health, the patient was diagnosed with Mycobacterium tuberculosis (MTB) infection and received Antitubercular medication. In 2021, stomach complications arose, leading to the diagnosis of Signet ring cell type stomach adenocarcinoma of stage IIB (TNM: $T_4N_0M_x$). Following a subtotal gastrectomy, she was given 6 cycles of chemotherapy regimen involving 5-Fluorouracil with Leucovorin and Oxaliplatin.

6.2 Common Cancer associated Genetic Variants

The patient's sample revealed the presence of germline mutations in *CDH1*, *BRCA1*, and *BRCA2*. Additionally, a non-synonymous variant Chr19(GRCh38):g.10987671A>G (dbSNP: rs750547893) was identified in *SMARCA4* gene in heterozygous condition and was identified as a rare variant of uncertain clinical significance (Table 6.1).

Within the exonic regions of *MUC3A*, a total of 80 non-synonymous variants were detected, with 19 being novel and 61 previously reported in the dbSNP database. Furthermore, considering the patient's MTB positive status, genetic variants associated

with MTB susceptibility were screened, revealing one variant Chr2(GRCh38):g.230185999A>G (rs3948464) of SP110 in a homozygous condition.

6.3 Tuberculosis Susceptibility genetic Variants

The variant rs3948464 identified in the patient's sample, located in *SP110*, has been previously noted for its significant association with tuberculosis (TB) in multiple populations, including three populations of western Africa and the Taiwanese population, in both active and latent forms (Sotudeh et al., 2019; Su et al., 2022). Despite the patient not displaying distinct symptoms of TB, such as cough, weight loss, or fever, only a pulmonary nodule was observed in the right lung. This suggests a potential connection between the *SP110* variant and the presence of a pulmonary nodule, warranting further investigation into its implications in the context of the patient's health.

6.4 Presence of rare variant of *SMARCA4*

Approximately 5 years later of breast cancer diagnosis, the patient was again diagnosed with intestinal-type gastric cancer (GC). Whole exome sequencing (WES) revealed the presence of a rare variant (rs750547893) in the gene *SMARCA4*.

6.5 *MUC3A* Variants

Comparison of the additional whole-exome sequencing (WES) data from gastric cancer (GC) patients and healthy controls revealed that, in the specific case of the discussed gastric cancer, 13 out of 19 novel variants and 35 out of 61 reported variants of *MUC3A* were uniquely identified. This underscores their exclusivity to this particular instance. Among the 35 reported variants found exclusively in the case, 4 were present in exon one, 1 variant each in exons six, seven, and eight, and 28 variants in exon two. Pathogenicity prediction for the 35 variants indicated tolerance, although Polyphen2_HDIV predicted the variants "Chr7(GRCh38):g.100959644C>A" (rs73398734) and "Chr7(GRCh38):g.100959664G>T" (rs73398735) as pathogenic. Further investigation is necessary to understand the potential implications of these variants in the context of gastric cancer and its progression.

Table 6.1: Variations in the *CDH1, BRCA1, BRCA2*, and *SMARCA4* genes observed in a patient with a history of Breast Cancer and later diagnosed with Gastric Cancer.

Variant	Variant Type	dbSNP ID	Disease	Supporting Read	Clinical Significance
CDH1					
Ala692Ala	Synonyous	rs1801552	• Hereditary diffuse gastric cancer • Hereditary cancer-predisposing syndrome	104	Benign
BRCA2					
Asn372His	Non-Synonyous	rs144848	• Breast carcinoma • Hereditary breast and ovarian cancer syndrome • Fanconi anaemia • Breast-ovarian cancer • Hereditary cancer-predisposing syndrome • Familial cancer of breast • Ductal breast carcinoma	79	Benign
Val2466Ala		rs169547	• Hereditary breast and ovarian cancer syndrome • Breast and Ovarian cancer • Ductal breast carcinoma	43	
BRCA1					
Ser1634Gly	Non-Synonyous	rs1799966	• Breast carcinoma • Cancer of Breast and Ovarian (Hereditary) • Malignant tumor of breast • Breast-ovarian cancer (familial 1) • Pancreatic cancer • Cancer-predisposing syndrome (Hereditary) • cancer of breast (Familial)	105	Benign
Lys1183Arg		rs16942		102	
Glu1038Gly		rs16941		97	
Pro871Leu		rs799917		90	
SMARCA4 Met289Val	Non-Synonyous	rs750547893	• Rhabdoid tumor predisposition syndrome 2 • Hereditary cancer-predisposing syndrome	85	VUS

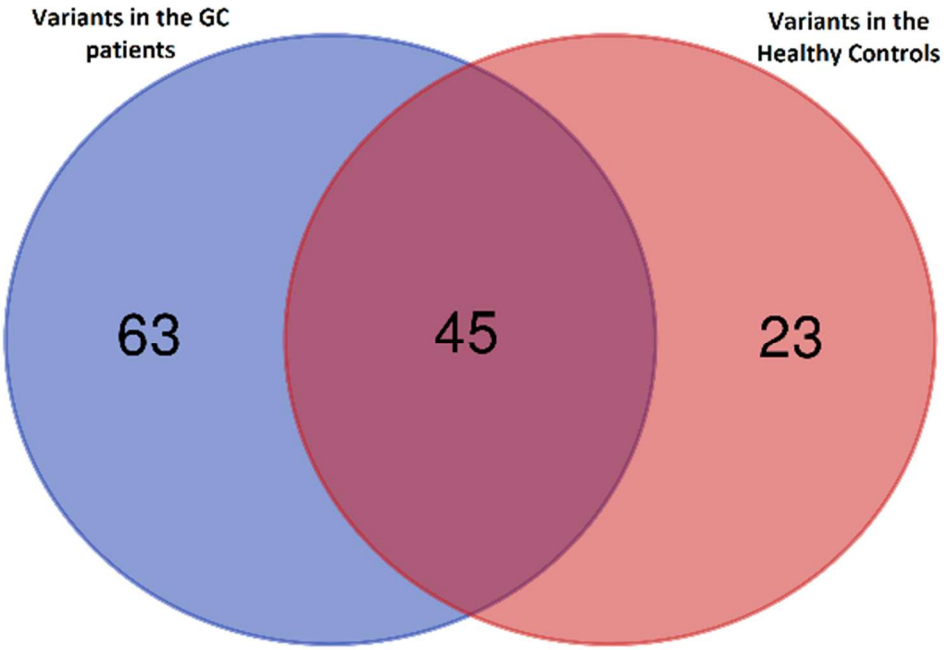
The important mucus gel protein-coding gene, *MUC3A*, has been strongly linked to a number of cancer forms., as reported in studies by Sotoudeh et al. (2019) and Su et al. (2022). Alterations in mucin expressions have been observed in both early and late stages of cancer, according to King et al. (2017). While *MUC3A* gene mutations were less prevalent in breast and gastric cancer samples, a detailed analysis using cBioPortal indicated a mutation occurrence of 2.04% in gastric cancer and 0.26% and 0.24% in two breast cancer datasets. The presence of a large number of SNPs within a single gene is quite uncommon. Therefore, 48 variants exclusively found in gastric cancer patients were investigated for their functional implications using Expression Quantitative Trait Loci (eQTL) analysis in the Genotype-Tissue Expression (GTEx) portal.

Despite the absence of the variants identified in the GTEx portal's tissue datasets, the uncertain pathogenicity of these novel variants introduces the intriguing possibility that they might contribute to the development of multiple cancers in the patient. The lack of available computational predictions of pathogenicity and literature data makes it challenging to categorize these variants definitively. Therefore, these variants are likely to be classified as Variants of Uncertain Significance (VUS). In a broader investigation, the prevalence of *MUC3A* variants was assessedby including the presented case along with 21 additional GC patients and 27 healthy participants from the same population,. The findings showed that cancer patients had more than twice as many variations as healthy controls. One-way ANOVA analysis revealed a significant difference between mean number of *MUC3A* variations of the two groups. Specifically, the gastric cancer group exhibited a markedly higher mean number of variants. This significant discrepancy suggests a potential association between the genetic profile of *MUC3A* and the susceptibility to gastric cancer in this population. The findings underscore the importance of considering the genetic variations in *MUC3A* as potential contributors to the increased risk of gastric cancer, warranting further investigation into the functional implications of these variants in the context of cancer development.

Further analysis using Fisher's exact test for each *MUC3A* non-synonymous variant occurrence revealed that out of 108 missense variants in gastric cancer patients,

70 had a significant association with cancer. Among these, 20 were exclusively present in the case, and notably, 9 of them were considered novel variants without any assigned dbSNP IDs. In the current example, there is a possibility that the rare variant rs750547893 and a significant number of mutations in the NM_005960 transcript (exon 2) of *MUC3A* could have contributed to the development of breast cancer and, later, gastric cancer, suggesting a potential association between these specific variants and gastric cancer.

A



B

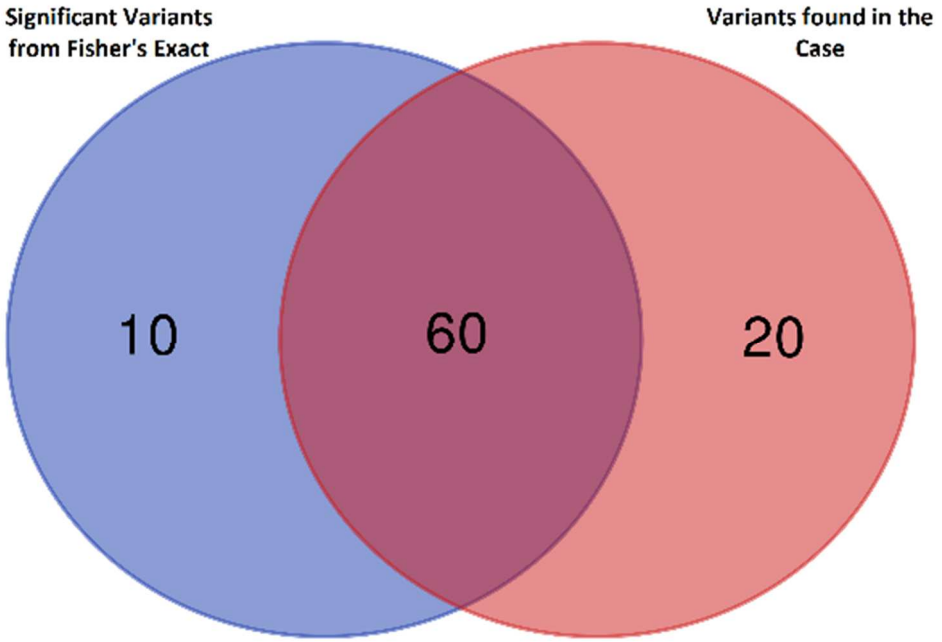


Figure 6.1: Number of *MUC3A* variants: **A.** Analysis of Missense Variants in Gastric Cancer Patients vs. Healthy Controls: In this examination, the focus was on assessing the distribution of missense variants in *MUC3A* between two groups - specifically, gastric cancer patients (n = 22, which includes the case) and a control group of healthy individuals (n = 27). **B.** Contrast in MUC3A Variant Numbers: Case vs. Fisher's Exact Test: This aspect involved a comparison between the case and the set of 70 variants identified through Fisher's exact test as having significant associations with cancer.

This incidental finding along with all the supplementary data is published and is available as:

"Sarma, R.J., *et al.*, (2023). Novel germline variants of *MUC3A* in a patient with ER+ breast cancer and signet-ring cell stomach adenocarcinoma**. Gene Reports**, 33, 101803. https://doi.org/10.1016/j.genrep.2023.101803."

CHAPTER 7
SUMMARY

The major findings of the study are:

1. The study explored initially healthy genomes and healthy exome from North-east india and exclusively from Mizoram. A total of 67 Very Important Pharmacogenes were screened in the whole dataset which led us to understand the allele frequency difference in the Mizoram population compared to NE population.
2. The study also led to the development of **MPVardb**, a database of variants of pharmacogenomic importance and their distribution among the population sets.
3. From the healthy exome study, there were 490 variants within the coding, 1112 in intronic, 95 in 3'UTR, and 52 in 5'UTR regions. Among the variants, 163 variants were non-synonymous, and 315 variants were found to be synonymous. Moreover, 1 frame-shift substitution, 4 start-loss, 1 stop-loss, and 2 stop-gain (2) variants were found.
4. About 77,799 variants were found in 67 VIPs in the NE IndiGenomes dataset.There were 2527 variants in the exonic region, whereas 67,144 variants were found in intronic regions. There were 2086 variants in 3'UTR3, 378 variants in 5'UTR, 107 variants in intergenic, 82 variants in ncRNA_exonic, 5445 variants in ncRNA_intronic, 24 variants in splicing, 5 variants in upstream, and 1 variant in exonic splicing region. Among the other variant types, 910 variants were synonymous, and 1462 variants were non-synonymous
5. This is the first work done for cataloguing the variants from northeastern populations, primarily the Mizo population as the region is known to be genetically diverse.
6. The number of variants in MPVardb is expected to grow in the future upon recruiting more volunteers from the region to understand the pharmacogenomic landscape to personalize therapies.

7. PharmCat annotation tools were used to screen clinically actionable (CA) pharmacogenomics (PGx) variants, based on PharmaGKB-DPWG and CPIC guidelines.
8. Healthy exomes (n=27) were analyzed to study the prevalence of CA PGx variants with different levels of function (decreased, intermediate, variable).
9. *DPYD* star alleles were also investigated in the dataset in the Healthy exome dataset.
10. A variant (Chr4:88131171 G>T or rs2231142) in *ABCG2* with a decreased function for allopurinol and rosuvastatin was observed, with allele frequencies of 0.259 in healthy exomes and 0.811 in IndiGen NE datasets.
11. Another variant (Chr19:40991369 C>T or rs8192709) in *CYP2B6* with intermediate function for Efavirenz and Sertraline was detected, with allele frequencies of 0.111 in healthy exomes and 0.913 in IndiGen NE datasets.
12. A *TPMT* variant (Chr6:18130687 T>C or rs1142345) was found in one donor with an allele frequency of 0.01, making them an intermediate metabolizer for thiopurine-based drugs.
13. A variant (ChrX:154534495 C>T or rs137852314) in *G6PD* was identified, responsible for variable function in the metabolism of several drugs.
14. The *UGT1A1*6* variant (Chr2:233760498 G>A or rs4148323) was detected, affecting the metabolism of Atazanavir and Irinotecan.
15. A *SLCO1B1*14* variant (Chr12:21176804 A>G or rs2306283) potentially leads to decreased statin metabolism.
16. Two DPYD variants (rs1801159 and rs17376848) were identified, but no star alleles were assigned to rs17376848, and there were no clinical guidelines available for these DPYD variants.
17. Important clinically relevant variants in genes *ABCG2, TPMT, G6PD, UGTA1* and *SLCO1B1* were found with guidelines set up by either PharmGKB-DPWG or CPIC.
18. In a study of 37 gastric cancer patients who experienced adverse drug reactions (ADRs) after chemotherapy, the majority of patients were male (75.68%), and the most common ADR was loss of appetite observed in 26 patients.

19. Prominent ADRs of different grades included fatigue (21 patients), nausea (23 patients), drowsiness (18 patients), dermatological adverse reactions (11 patients), vomiting (8 patients), and alopecia (characterized by hair loss) in 8 patients.
20. Concurrent ADRs were observed in many patients, with loss of appetite often co-occurring with fatigue, nausea, and drowsiness.
21. Among the 37 followed-up patients, exome sequencing and follow-up data were available for 25 patients.
22. The chemotherapy regimen primarily consisted of 5FU with leucovorin and platinum based drugs, administered intravenously to 19 patients and orally to 4 patients. A smaller number of patients received alternative chemotherapy agents.
23. In the CBC profiling after chemotherapy, neutropenia was observed in 3 patients, leukopenia in 2 patients, and low hemoglobin in 1 patient.
24. Overall, this study provides insights into the prevalence of ADRs in gastric cancer patients after chemotherapy, with a focus on the types of ADRs and the gender distribution among the patients.
25. *DPYD* variants, including rs1801159 (I543V), were identified in 35% of patients and were associated with DPD deficiency or toxicity. ADRs included loss of appetite, nausea, darkening of nails, severe neutropenia, and alopecia.
26. The *DPYD* variant rs1801159 was predicted to be damaging by multiple computational tools, including SIFT, Polyphen, and Mutation Taster.
27. *DPYS* variants, such as rs2298840 (F72F), were found in 45.76% of patients and associated with Dihydropyrimidinase deficiency. ADRs, including dermatological adverse reactions, were observed.
28. UPB1 variants, including rs2070475, were identified in 45.15% of patients and associated with beta ureidopropionase deficiency. ADRs, including mucositis, neutropenia, and diabetes, were observed in some patients.
29. Some patients had variants in multiple genes, *DPYD* and *DPYS* variants and experienced a range of ADRs.

30. eQTL analysis was conducted on variants of *DPYD, DPYS, UPB1*, and *TYMS*. This analysis revealed significant associations between certain variants and gene expression in various tissues.
31. *DPYD* upregulation was significantly associated with better survival, while *UPB1* upregulation was linked to poor survival but it was not statistically significant. This suggest that *UPB1* upregulation might have role in poor survival but it needs further analysis.
32. Analysis of The Cancer Genome Atlas (TCGA) data revealed mutation frequencies for *DPYD, DPYS, UPB1,* and *TYMS* in gastric cancer patients. *DPYD* mutations were more common compared to the other genes.
33. While analysing TCGA data, patients with *DPYD* and *DPYS* mutations appeared to have lower survival rates, however it was not found statistically significant.
34. Some patients had variants associated with 5FU, Leucovorin and Oxaliplatin (FLO) toxicity, such as rs13181 in *ERCC2*, rs17376848 in *DPYD*, and rs1801133 in *MTHFR*. These variants were present in high percentages of the patients.
35. Some of these variants were categorized as benign or of unknown significance, and their pathogenicity remains to be fully understood which need further validation.
36. Multiple variants responsible for FLO toxicity also co-occured in patients, that might have potential influence on the responses to treatment and survival outcomes. The presence of certain combinations of variants was observed in different patients.
37. The study found that the prevalence of various variants, such as rs13181, rs25487, rs1695, rs717620, and rs11615, was high among the gastric cancer patients. However, not all of these variants were found together in any single patient.
38. The presence of multiple variants in patients, especially three-star variants, suggests that complex genetic interactions may play a role in determining the outcomes of gastric cancer treatment.

39. The co-occurrence of some variants, such as rs1801133, rs25487, and rs11615, was observed in 10% of patients. This suggests that certain combinations of variants might have a cumulative effect on treatment outcomes.
40. While some patients with *DPYD* and *DPYS* mutations appeared to have lower survival rates, the findings were not statistically significant. This highlights the complexity of factors influencing patient outcomes on chemotherapy.
41. Patients exhibited variants in multiple genes simultaneously, potentially leading to a more complex genetic landscape and affecting their responses to chemotherapy that necessisated further in-depth study.
42. The study integrated data from various sources, including genetic variants, gene expression, survival outcomes, and mutation frequencies, to gain a comprehensive understanding of the genetic factors impacting gastric cancer treatment.
43. Survival analysis which was performed based on gene expression data, and it was revealed that *DPYD* upregulation had association with a lower risk (HR = 0.54) and better survival, while *UPB1* upregulation was found to be associated with a higher risk (HR = 1.47), suggesting poorer survival.
44. The study underscores the importance of considering genetic variants and their interactions in assessing the effectiveness and outcomes of adjuvant chemotherapy for gastric cancer patients.
45. Understanding the genetic factors influencing treatment responses may lead to the development of more personalized and targeted therapies for gastric cancer patients in the future.
46. The current finding of the study further suggest to unravel the complex genetic underpinnings of treatment responses in gastric cancer patients and potentially improve patient outcomes.
47. The case study performed on the patient with ER^+ breast cancer and Stomach Adenocarcinoma found statistically significant association of *MUC3A* variants with stomach cancer development within the Mizo patients and can put the population susceptible to stomach cancer (Sarma et al., 2023)

Appendix 1

Questionnaire for recruitment of volunteers for sequencing the Healthy Genome and Healthy Exomes

IndiGen: Mizoram University and Civil Hospital, Aizawl

Name (Hming): _____ ☐

Male Female

Marital Status (Nupui/Pasal nei/neilo):_____ Date of Birth (Pian ni):_____ (dd/mm/yy)

Corresponding address:_____

_____

Permanent address:_____

_____

Mob: _____Mob (Alt):_____

NAME OF RELATIVE/NEXT OF KIN

Father _____ ____ Phone no. _____

Mother _____ Phone no. _____

Spouse _____ Phone no. _____

LIFESTYLE HABITS

A. TOBACCO IN THE FORM OF SMOKING (MEIZIAL)

1) Do you smoke? (Meizial I zungaiem?) Yes ☐
   No

2) How old were you when you started smoking? (Kum engzat I nihin nge I zuk tan?)_____

3) If quit, since when?(I nghei tawh a nih chuan engtik atangin?)
   _____

4) What brand of cigarette do you usually smoke? (Eng meizial brand nge I zuk tam ber?)
   Zozial ☐ Branded ☐ Both ☐

5) How many sticks do you smoke per day? (Ni khatah tlawn eng zat nge I zuk?)
   :_____

B. TOBACCO IN THE FORM OF SNUFF (Hmuam chi)/ TOBACCO-SMOKE INFUSED WATER (Tuibur)

1) Do you take tobacco in the form of Tobacco smoke-infused water and/or snuff? (Tuibur/Sahdah I hmuam ngai em?) Yes ☐ No ☐

2) What type do you take? (Eng ang chi nge I tih thin?)
Sahdah ☐ Tuibur ☐ Both ☐

3) What type of snuff do you usually take? (Eng ang sahdah nge I hmuam thin?)
Khaini ☐ Raja ☐ Local ☐

How old were you when you started taking? (Kum engzat I nihin nge I tih tan?)

4) Sahdah/Khaini_____ Tuibur_____

5) If quit, since when? (I nghei tawh a nih chuan engtik atangin?)
Sahdah/Khaini_____ Tuibur_____

6) How many pinches of snuff do you typically use per day?
10 or less
(Ni khatah hmuam engzat nge sahdah I hmuam tlangpui thin?)
10-20

20 or more

7)) How much of tuibur do you typically use per day? 1 Koinonia bottle (150ml) or less
(Ni khatah tuibur engzat nge I hmuam thin 160ml to 300ml (=2x150 ml koinonia)
More than 300ml

C. TOBACCO INGESTED GUTKHA PRODUCTS (EI CHI)

1) Do you take any other tobacco products (Gutkha)? (Vaihlo atanga siamthil dang tih I nei em?)
Yes ☐ No ☐

2) What kind do you take? (Eng ang chi nge I ei thin?)
Shikhar ☐ Zarda pan ☐ Kuhva ☐
Specify _____

3) How old were you when you started taking gutkha products? _____
(Kum engzat I nihin nge I gutkha I ei tan?)

4) If quit, since when?(I nghei tawh a nih chuan engtik atangin?)
_____

5) How many pouches of snuff do you typically use per day?
1 or less
(Ni khatah fun engzat nge I ei tlangpui thin?)
2-4

5 or more

D. ALCOHOL CONSUMPTION

1) Do you take alcohol? (Zu I in ngai em?) Yes ☐ No ☐ Nghei (Quit ☐ Engtik (When):_____yrs

2) How old were you when you started drinking alcohol? (Kum engzat I nihin nge zu I in thin?)_____

3) Which type of alcohol do you mostly drink?(Eng zu nge I in tlangpui thin?)
Branded ☐ Local ☐ Both ☐

4) What kind of alcoholic beverage do you drink mostly?(Engang chi nge I in tlangpuiber?)
Whiskey ☐ Beer ☐ Rum ☐ Wine ☐ Vodka ☐

5) How often do you drink alcohol? 2 or less days a week
(Karkhatahni 2 aiatlem)
(Engtia zingin nge I in thin?) 3-4 days a week
(Karkhatahni 3-4)
More than 5 days (Ni 5 aia tam karkhatah)

6) How many pegs do you normally drink? 2-3 pegs ☐ 4-5 pegs
(Peg engzat nge I in thin?) 6-7 pegs ☐ More than 8 pegs

E. TASTE PREFERENCES (Ei leh in)

Do you consume? (I ei ngai em?)	0 (Never)	1 (Little) 1 days / week	2 (Average) 2-4 days / week	3 (Heavy) 5-7 days / week
Spicy food (Thil thak)				
Smoked Meat (Sa rep)				
Smoked vegetables (Thlai rep)				
Fried food (Chawhhmeh kan)				
Fruits (Thei)				
Boiled Mix Vegetable (Bai)				
Fermented Pork Fat (Saum)				
Canned Foods (A tin nei ang chi)				
Frozen Foods				
Soda drinks (Coke, Pepsi, etc)				
Fruit drinks (Thei tui)				
Tea : Milk ☐ Sugar ☐				
Water (Tui)				

Do you re-use oil for cooking/ frying? Yes ☐ No ☐
(Chawhmeh kana tel hman tawh hnu in hmang nawn thin em?)

F. ENVIRONMENTAL FACTORS

1) Is there a cell phone tower near your house or workplace? Yes ☐ ☐
(I chena emaw I thawhna hmunah emaw cell phone tower a awm hnai em?)

2) Are you exposed to jhum cultivation? (Lo halna hmun I hnaih em?) Yes ☐ No ☐

3) Does your work involve exposure to sunlight? Yes ☐ No ☐
(Ni sa do ngaihna hna I thawk em?)

2-3 hours daily

(Nitindarkar 2-3)

4-5 hours daily

(Nitin darker 4-m More than 6 hours (Darkar 6 aiareinitin)

4) Do you use Cosmetics? (Cosmetics I hmang ngai em?) Regularly ☐ Occasionally ☐

5) Are you exposed to secondary smoking at home or at your workplace?
(I chenna hmunah emaw I thawhna hmunah emaw I bul a miten meizial anzu em?

Everyday (Nitin) ☐ Occassionally (A changzeuhzeuh)

6) Was your mother smoking when you she was pregnant with you? Yes ☐ No ☐ Don't know ☐
(I nuin a pailai che in mei a zu thin em?) Everyday [Nitin] ☐
Occassionally (Zu zeuhzeuh)

7) How often do you exercise (Exercise I la ngun em?)_____

8) Do you chew betel nut? (Kuhva I ei ngai em), I yes, how much per day:_____

9) Do you work in any of the following? (Heng hmunah hian I thawk em?)

Place of Work	How many years? (Kumengzat?)
Quarry	
Automobile workshops	
Office work	
Animal Husbandry	
Agriculture	
Driver	
Carpentry	
Construction worker	
Tuibur/Vaihlo factory	
Quarry	

HEALTH AND FAMILY INFORMATION

1) Do you have any allergies? (Allergy I nei em?) Yes ☐ No ☐

If yes, specify (A awm chuan min lo thai lan sak):

_____

2) List out your family member place of origin and sub-tribes

Relation	Place of origin	Sub -Tribe (Hnam)
Father (Pa)		
Mother (Nu)		
Paternal Grand Father (Pa – pa)		
Paternal Grand Mother (Pa – nu)		
Paternal Great Grand Father (Pa – pu)		
Paternal Great Grand Mother (Pa – nu)		
Maternal Grand Father (Pa – pa)		
Maternal Grand Mother (Pa – nu)		
Maternal Great Grand Father (Pa – pu)		
Maternal Great Grand Mother (Pa – nu)		

4) What language do you use in the house (Eng tawng nge inah in hman)?

_____

CONSENT (*REMTIHNA*)

Heng achunga thute hi ka hriatpui in, ka thisen hi zirchian atan pek ka remti thlap e.(*The information provided above was given with my full consent and I do not have any objection in providing my biological sample for research purposes. I have read and understood the consent information.*

Hmun*(Plac*e): Signature:

Date: Hming (*Name)*:

Appendix 2

Questionnaire for recruitement of Gastric Cancer Ptients

Questionnaire for Epidemiological Study of Gastric Cancer

Referring Dr:_____ MSCI/Civil Hospital No._____/_____

Referring Unit:_____ Reg Date:_____

PROFORMA

MZU,MSCI,CIVIL Hospital & NIBMG

PERSONAL HISTORY

Hming (*Name)*: Mipa/Hmeichhia (*Male/Female*):

Kum (*Age):*

Tawng hman (*Language*): Nupui/pasal nei/neilo (*Marital status):* Pian ni*(Date of birth):* *Nupui/pasal neiha kum zat* (*Age at the time of marriage*):

Rihzawng (Weight): San zawng (*Height*):

Lehkha zir chen(*Education*): Eizawnna (*Occupation)*:

Unau engzat nge in nih*? (No. of Siblings):* [ ] Mipa (*Male*) [ ] Hmeichhia (*Female*) [ ]

Fa I nei em? (*Do you have children?):* Aw/*Yes* [ ] Aih/*No* [ ]

I neih chuan, fa engzat nge I neih? (*If yes, how many children do you have?):* [ ]

Mipa/Hmeichhia engzat nge? (*Gender of the children):* Mipa(*Male)* [ ] Hmeichhia(*Female)* [ ]

(Thi sa a piang chhiar tel tur, chhiat erawh chhiar tel loh tur) (*Please include stillbirths; it is not necessary to include miscarriages)*

Addres:_____ _____ _____Pin Code_____

Tel No._____Mob.No._____ _____

E mail:_____

Cancer
Diaognosis/Treatment_____
____

Engtik kumah nge cancer I vei tih hmuhchhuah a nih? *(Year of cancer detected?):*_____

Tumor	Site	Age	Histopathology	Surgery Date	Chemotherapy Date	Radiation
1st Primary						
2nd Primary						
3rd Primary						

Syndrome Diagnosis:

Consent: Yes/No
Date:_____

Blood collected: Yes/No Date:_____ Received on_____
From_____

Second sample collected: Yes/No Date:_____Received by_____Thru_____

Tumor Tissue Collected: Yes/No Date_____ICG/Other Biorepositary_____

Details taken by:_____
Date:_____

Genetic Pre-Test Counseling done by :_____Date:_____

Genetic Post-Test Counseling don by :_____Date:_____

FAMILY INFORMATION:

In chhungkua ah natna dang vei in awm em(cancer ni lo)*(Any other type of diseases in the family (other than cancer)*:

Name Relation Education Sex Age Disease Information Occupation Habit Signature

In chhungkua ah Cancer vei dang an awm em *(Does anyone else in your family have cancer)*:

Name Relation Education Sex Age Disease Information Occupation Habit Signature

Hereditary: Yes[ ] No[ ] Autosomal Dominant: Yes[ ] No[ ]
Autosomal Recessive: Yes[ ] No[ ] Sex linked: Yes[ ] No[ ]
[ ] Cannot ascertain/Not applicable
[ ] Sporadic [ ] Early Onset [ ] Routine RET [ ] Familial
[ ] Others_____

Chhungkaw member zat (*Number of family member*):
a) Tunah (*Now*): Puitling(*Adult*) - Mipa(*Male*): [ ]; Hmeichhia(*Female)*: [ ];

Naupang (*Children)* – Mipa(*Male)*: [ ]; Hmeichhia(*Female)*: [ ]

b) Boral tawh (*Decease number*): [ ]
Boral chhan (Reason) – Pumpui cancer (*Gastric cancer*): [ ]; Adang (*Others*): [ ]

MSCI/Civil hospital No._____

PEDIGREE

(Draw pedigree one degree above and below affected individuals and note consanguinity.)

GEOETHNIC ORIGIN

	Place of birth (Dist./State)	Present place of stay (Dist./State) & duration	Dist./State of origin	Family name/Surname	Religion	Tribe	Sub tribe
Index							
Father							
Mother							
Paternal Grandfather							
Paternal Grandmother							
Maternal Grandfather							
Maternal Grandmother							
Remarks							

Environmental/Lifestyle Factors

What has been your main occupation?_____

Hengah te hian hna I thawk em? I hnathawhnaah hetiang te hi I in chiahpiah tir em? *(Do you have Occupational exposure to?)*		No. of years	Age (From/to)	Nature of use	Name of company/brand
Radiation(eg. In a factory,laboratory/medical setting)	Yes No Don't Know				
Plastic	Yes No Don't Know				
Agriculture/Rubber plant (If yes C4A)	Yes No Don't Know				
Pesticides/Pest control/ Mosquito Repellant	Yes No Don't Know				
Chemical/Dyes	Yes No Don't Know				
Any other exposure (Asbestors,Chromium or Lead)	Yes No Don't Know				

i) Was your mother an agriculture worker around the time of your birth? Yes/No

ii) Has DDT ever been used in or around your household? Yes/No

iii) What is your water supply source? River [ ] Tube well[ ] Govt./municipal [ ]

iv) Other_____

I hna a hahthlak viau em, zan lam ah hna I thawk em(night duty)? (*Is your job stressful or do you perform shift work (night duty)?)*: Aw/*Yes* [ ] Aih/*No* [ ]

In in bulah cell phone tower a awm em?*(Is there a cell phone tower near your house?)*:

Aw/Yes[ ] Aih/No [ ]

TASTE PREFERENCES:

Do you consume (I ei ngai em)	0(Never)	1(Little) 1 days in a week	2(Average) 2-4 days in a week	3(Heavy) 5-7 days in a week
Spicy food				
Western food (Pizza,burgers,fries)				
Burmies product				
Sour test (tamarind ,lime juice etc)				
Bawngsa (*Beef*				
Vawksa (*Pork)*				
Kelsa (*Mutton)*				
Arsa (*Chicken)*				
Artui (*Egg)*				
Sangha (*Fish)*				
fermented fish				
Bekang/fermented pulse				
Sa-Um				
Extra salt with food				
Pickles/chutneys				
Smoked vegetables				
Smoked meat				
Fat intake				
Boiled food				
Fried food				

Smoked food				
Fibers food/fruits (Banana				
Azinomoto				
Soda(sodium by carbonate)				
Vinegar				
Salt(packed / raw)				
Oil (				

A tlangpuiin I chaw/chawhmeh te I ei thin dan? *(How do you normally consume your food items?)*: Chhum */Boiled* [ ]; Kan/F*ried* [ ] Smoked [ ]

Tobacco & alcohol History:

Zu I in em? (*Do you consume alcohol?):* Aw/*Yes* [ ] Aih/*No* [ ]

Aw (*If yes*): In reng (*Regularly*) [ ] A chang chang in (*Occasionally*) [ ]

Engtik atangin nge I in tan? (*When did you start taking alcohol?)* :

Eng zu nge I in? (*Type of alcohol)*: Mizo siam *(Local)* [ ]; Hnamdang siam (*Branded)* [ ];

A pahnih in (*Both)* [ ]

I nghei tawh anih chuan, engtik atangin? *(If quit already, since when?)*:

Have you ever consumed any alcoholic beverages, such as wine, beer or spirits at least once a week for six months or longer?

[ ] Yes [ ] No [ ] Don't know

If consumption has changed during life record highest consumption.

Beverage	Yes/No	From age	To age	Units/Day	Days/Week

Spirit shot = 50 ml, 1 bottle = 15-20 shots

Tuibur I hmuam em? (*Do you consume tuibur?):* Aw/Yes [ ] Aih/No [ ]

Aw *(If yes)*: Hmuam reng (*Regularly)* [ ] A chang chang in (*Occasionally)* [ ]

Engtik atangin nge I hmuam tan? (*When did you start taking tuibur?)*:

Eng ang tuibur nge I hmuam? (*Type of tuibur):* Bazar a lei (*Local)* [ ] Mahni a siam (*Self-made)* [ ]

I nghei tawh anih chuan, engtik atangin? (*If quit already, since when?):*

If consumption has changed during life record highest consumption.

Beverage	Yes/No	From age	To age	Units/Day	Days/Week

Mei I zu em? (*Do you smoke?):* Aw/*Yes* [ ] Aih/*No* [ ]

Aw (*If yes*)*:* In reng (*Regularly*) [ ] A chang chang in (*Occasionally*) [ ]

Nikhat ah engzat nge I zuk thin tlangpui *(Average Number of smoke per day)*:

Engtik atangin nge I zuk tan? (*When did you start smoking?)*:

Eng nge I zuk? (*Type of Smoke)*:Zozial*(Local)* [ ]; Biri [ ];Cigarette(Eng siam?)*(Which brand?):*

I nghei tawh anih chuan, engtik atangin? *(If quit already, since when?)*:

Has there ever been a time when you smoked at least one cigarette per day for three months or longer?

[ ] Yes [ ] No [ ] Don't know

If yes list consumption (excluding times when the subject did not smoke)

Product	Yes/No	Used From/To	Frequency	Av. Quantity per day
Cigarette				
Biri				
Zozial				

Vaihlo a siam thil dang tih I nei em? (*Do you consume other tobacco products?):*Aw/Yes [ ] Aih/No [ ]

Aw *(If yes):* Ti reng *(Regularly)* [ ] A chang chang in(O*ccasionally)* [ ]

Have you ever chewed pan or tobacco regularly?(At least once a week for six months or more)Yes [ ] No [ ] Don't Know [ ]

Type	Yes/No	From age	To age	No. per day
Chewing with tobacco and lime(khaini) Pan+tabacco+betelnut+lime+catechu(mewa)				
Gutka				
Sahdah(*Oral snuff*)				
Kuhva*(Pan/Beetle nut)*				
Zarda Pan				
Supari				
Chewing without tobacco (eg. pan without tobacco)				
Adangte *(Others)*				

History of passive smoking:

Do any of your family member/colleagues smoke tobacco at home? Yes/No

Medical History:

I blood group eng nge? *(What is your Blood group?)*

A+ [ ] A- [ ] B+ [ ] B- [ ] AB+ [ ] AB- [ ] O+ [ ] O- [ ]

Exercise I la ngai em? *How often do you exercise?*

Ngai lo(N*ever)* [ ]; Karkhatah vawi khat aia tlem(*Less than once a week)* [ ];

Karkhatah vawi khat(*Once a week)* [ ]; Karkhatah vawi 2-3 (*2-3 times a week)* [ ];

Karkhatah vawi 4-6 (*4-6 times a week)* [ ]; Nitin(*Everyday)* [ ]

Ultrasonography:

Other: Region_____Report

Date:_____Impression_____

CT scan: Region_____Report Date_____

impression_____

_____

_____

Colonoscopy/Endoscopy:

Regions_____Date_____

Impression_____

_____

Natna/Damlohna dang I nei em? (*Do you have any other diseases?):* Aw/*Yes* [ ]

Aih/*No* [ ]

I neih chuan, eng natna nge? (*If yes, what type of disease?):*

_____

H. pylori [ ] Diabetes [ ] obesity [ ] HIV [ ] HbsAg[ ]

HCV[ ] EBV [ ] Gastric atrophy [ ]

Surgery:

Site/Procedure_____

Pathological Staging-

pTNM_____Date_____

Histopathological Report: Specmen_____Path No._____

Date_____Impression_____

_____

_____

IHC: Hormone receptor status

Tumor details: Specimen_____Path No._____

Report Date_____Grade_____Size of the tumor_____

cm. Tumor emboli_____Lymphovascular Invasion_____

Syndromic features noted:

A hnuai ami te hmang hian enkawl I ni tawh em? History of taking HRT/Reflux /Proton Pump Inhibitors/ Others(Give details)_____

_____

___

Obestric History:

Gravity/Parity

Recurrent spontaneous abortions

Still births/ Neonatal deaths

Congenital malformations

Others

Remtihna *(Consent):*

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

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

Hmun*(Plac*e): Signature:

Date: Hming (*Name)*:

KA LAWM E

(THANK YOU VERY MUCH FOR YOUR HELP)

Follow- Up Notes

Appendix 3

Questionnaire for recruitement of Gastric Cancer Ptients for Pharmacovigilance

Adverse Drug Reaction Reporting (GASTRIC CANCER)

Mizoram University and MSCI

ABOUT THE STUDY

Adverse drug reaction (ADR) is a serious medical concern globally. The unusual and harmful effect of any drug at optimum dose is referred to as ADR. ADR could be seen in many disease treatments conditions and it contributes significantly in deteriorating the patient health. Anticancer drugs contribute significantly to the global burden of ADRs. Variants in the genes encoding drug metabolising enzymes may cause major structural change in the protein which may make the enzymes more or less effective. Impaired enzymes do not metabolise the drug efficiently and lead to increased concentrations of the medication which may lead to increase in the toxicity in the body. This study is aimed to investigate the genetic variation of the genes responsible for all forms of drug metabolism in relation to patient's response to the drug and also the genetic variants responsible for ADR associated with Gastric Cancer.

- Adverse drug reaction hi Mizo chuan damdawi ngeih lo kan tih ang hi ani a,hetiang dinhmun hi natna hrang hrangah hmuh tur a awm zel ani.
- Daktawr in damlote damdawi an chawh te an ngeih loh/huat palh a awmin damlo ten an dampui aiin an tlakchhiat phah thei thin.
- Heng zingah hian Cancer te tana damdawi an chawh te, chemotherapy an pekte hi damdawi ngeihloh palh awm thei zinga a larzual te an ni.
- Damdawi kan ei/lak hian a thawh tur a thawh theih nan kan taksa hian alo phelsawm thin a,a phelsawm tute hi an hmingah 'enzyme' kan ti thin.
- Heng enzyme te hian an hna thawk tur chuan,ruang am bik tak an neih angaia, chu an ruangam chu alo danglam chuan hna an thawk theiloa, damdawi pawhin a thawh tur a thawklo mai nilovin taksa tan a hlauhawm thei hial zawk a ni.
- Mi hrang hrangte kan DNA ah kan danglam avang hian, kan damdawi ngeih leh ngeihlo tur pawh a danglam thin a,he zirnaah hian Gastric cancer te hnena damdawi an chawh te a zirin Gastric cancer nei lai mekte tana an ngeih theih ber tur DNA atangin zir kan tum ani.

PROCEDURE

Our research staff/ project personnel will collect the data pertaining to your treatment file in MSCI/ Civil Hospital Aizawl and also get information from you about the medication, health status, life style and food habits.

(Research staff/Project personnel ten MSCI/Civil Hospital Aizawl a in file neihte atangin kan hriat duhte kan lakhawm anga, chuan damdawi ei/lakdan te,I hriselna dinhmun te, I nunphung leh ei leh in te pawh tiamin lak khawm a ni ang.)

CONSENT STATEMENT

- A member of the research team from MZU has informed me about the study and discussed with me the requirements for participation in this study.
  (MZU a research team ten he zirnaa a tel tur te tana pawimawh leh tul te min hrilh vek e.)
- I have read all of the information contained in this Information Sheet (or had it read to me), and I have had time to think about the information, and all of my questions have been answered to my satisfaction.
  (He remtihna form hi uluk takin ka chhiarchhuak veka/min chhiarchhuah sak a,ka hriatthiam lohte sawifiah in leh ka zawhna te ka duh ang thlapa min hrilhfiah hnu in,uluk taka inngaituahna hun ka nei e).
- I voluntarily agree to be part of this research study, to follow the study procedures, and to provide necessary information to the investigator or other staff members as requested.
  (He zirnaah hian ka remtihna ngeiin ka tel a,a procedure te zawmin,a zirtu ten ka laka an mamawh engkim chu ka pe ang.)
- I am under no pressure to participate in the study, and I understand that I may withdraw from the study at any time. I also understand that my participation in the study may be terminated by the study investigator if necessary.
  (He zirnaa tel tur hian hnawn luih emaw, tih luihna in ka tello a.Ka duh hun hunah ka in hnuk dawk thei ani tih ka hriat angin,a zirtu ten an duh hun hunah min ban thei ani tih ka hria e.)
- There are no risks to the patients participating in the study. All precautions will be taken while conducting the questionnaire data as it will be done by trained personnel with expertise. There is no financial cost of the study to the participants.
  (He zirchianna a damlo tel te tan hian hlauhawm a awm lova.Zawhna leh chhanna zawt tur a trained bikten fimkhur takin an zawt ang.He zirchianna a damlo tel te tan chawi ngai a awmlo ang.)
- If needed, the research staff involved in the study may contact you (for getting information on your health status/ life style and food habits) either in person or through your preferred means of communication – phone or email etc.
  (A tulna awm thei ah hmaichhan ah emaw phone call in emaw email hmangin he zirchianna a thawktu te atangin biak pawh in ni ang.)
- In the event of any reports or publications resulting from this study, no information will be revealed that will permit readers to identify you. The data will be made accessible to the scientific community, upon removal of all identifiable information. All the information obtained in this study will be kept confidential to the extent permitted by law and National Ethical guidelines for Biomedical and Health Research Involving Human Participants 2017. The research findings will be shared with the Clinicians in MSCI and Civil Hospital Aizawl and can be obtained from them, if interested.
  (He zirchianna a tel te nihna leh an result te tunge an nih tih hriat na tur awmthei angte thupsak an ni anga.MSCI leh Civil Hospital Aizawl a Clinicians te bul ah chiah hriattheih in a awm ang.)
- The data generated from the research study will be stored confidentially and kept in locked cabinets as well as password protected computers. The data will be used solely

only for research purpose and will be handled only by the Investigators, authorized research personnel's and project staff.
(He zirchianna a data kan lak ang ang te uluk taka dahthat a ni anga,investigator leh research team ten research hna atan chauh hman ani ang.)

Please give YES beside each statement you agree with

1. I agree to participate in the studies that will involve experiments using DNA isolated for Next Generation Sequencing of whole genome/Whole Exome/sanger sequencing from my whole blood after anonymization _____
(He zirnaah hian ka thisen atanga lak DNA hmanga test hrang hrang neih tur Next Generation sequencing hmanga Whole Genome emaw Sanger sequencing chu keimah,mimal taka min lo chhui let theihna awm lova hna an kalpui turah hian rem ka ti e.)

2. I agree to provide blood sample for biochemical investigation. _____
(May not be required in all cases)
(He zirnaah hian ka thisen,Biochemical test atan phal takin ka pe e.)

3. I agree for long term storage of biological samples and research data obtained through this project for future research _____
(He zirna atana ka thisen pek leh ka taksa ruangam tehna te, tuna an zirna atanga rahchhuah:data pawh hi nakin zela an zirchianna atan ka pe phal e.)

4. I agree in providing all my medical records that might be useful for this study_____
(He zirna atana ka medical records in tangkaina a neih theih chuan ka pe phal e.)

Contact Info of Investigators:

Name & Address: Prof. N. Senthil Kumar, Mizoram University, Department of Biotechnology, Aizawl,Mizoram ,796004 . Mobile: +91 9436352574 Email: nskmzu@gmail.com
Name & Address: Dr. Jeremy L.Pautu, HoD, Medical Oncology, MSCI, Zemabawk,Aizawl, Mizoram, 796017. Mobile: +919436150301 Email: jlpautu@gmail.com
Name & Address: Dr. B. Zothankima, HoD, Radiation Oncology, MSCI, Zemabawk, Aizawl, Mizoram, 796017. Mobile: +919436142185 Email: drzothana72@gmail.com
Name & Address: Dr.K. Lalfakzuala, Specialist, Radiation Oncology, MSCI, Zemabawk, Aizawl, Mizoram, 796017. Mobile: +919436151971 Email: valteak1@gmail.com

Name (BLOCK LETTER): _____SIGNATURE: _____

Date: Place:

Follow-up for ADR - Pharmacovigilance study (GASTRIC CANCER)

1. **Primary Details(Nihna kimchang):**

Patient ID	:	G	C	/						Date	:	
Patient Name	:									Gender	:	M F
Address	:									Age	:	
Ethnicity	:											
Telephone/Mobile	:											

2. **Pathology Details(Pathology kimchang) (To be filled up from earlier reports):**

Cancer Type (Subtype should be written below):

S. No	**Type**	**Subtype**	**TNM Staging**	**Symptoms**	**EBV Status**	***Helicobacter Pylori* Status**	**Gastritis status**

3. **(i) Medication Details(Damdawi kimchang):**

Medicine Name (Damdawi hming)	Administration route(Damdawi ei dan)	Number Times per Day(Nikhatah vawi engzah nge)	Doses per time(Dose engzah nge)	Start Date (Ei tan hun)	End Date (if stopped) (Ei tawp hun)	Reason of discontinuation (Ei chhun zawm loh chhan)

ii. Details of any medicines skipped (Damdawi skip kimchang):

Medicine Name (Damdawi hming)	How many Doses skipped (Dose engzah nge I skip)	Skipped Date (Skip ni)	Reason of skipping (Skip chhan)	Resume Date (If continued again) (Ei chhunzawm ni)

4. **i. Combinations of Drugs**:

Medicine Name(Damdawi hming)	Administration route(Damdawi ei dan)	Number Times per Day(Ni khatah vawi engzah nge I ei)	Doses per time(Dose engzah nge)	Start Date(Ei tan hun)	End Date (if stopped) (Ei tawp hun)	Reason of discontinuation (Ei chhunzawm loh chhan)

ii. Details of any drug combinations skipped:

Medicine Name(Damdawi hming)	How many Doses skipped (Dose engzah nge I skip)	Skipped Date (Skip ni)	Reason of skipping (Skip chhan)	Resume Date (If continued again) (Ei chhunzawm ni)

5. **i) Chemotherapeutic Drug:**

Medicine Name(Damdawi hming)	Administration route(Damdawi lak dan)	Number Times per Day(Ni khatah vawi engzah nge I lak)	Doses per time(Dose engzah nge)	Start Date(Lak tan hun)	End Date (if stopped) (Lak tawp hun)	Reason of discontinuation (Lak chhunzawm loh chhan)

ii) Radiotherapy:

Medicine Name(Damdawi hming)	Administration route(Damdawi lak dan)	Number Times per Day(Ni khatah vawi engzah nge I lak)	Doses per time(Dose engzah nge)	Start Date(Lak tan hun)	End Date (if stopped) (Lak tawp hun)	Reason of discontinuation(Lak chhunzawm loh chhan)

ii. Details of Chemotherapeutic Drug / Radiotherapy combinations skipped:

Medicine Name(Damdawi hming)	How many Doses skipped(Dose engzah nge I skip)	Skipped Date(Skip ni)	Reason of skipping(Skip chhan)	Resume Date (If continued again) (Lak chhunzawm ni)

6. Details of Adverse Reactions:

I. Basic Details(Side effect kimchang)::

a. Was there any side effect? (Side effect a awm em?) YES NO NOT NOW

b. If yes, When did side effect start? (A awm chuan engtik atang nge?)

_____

c. Is Side effect still continuing? (Side effect ala awm reng em?) Yes(Aw) / No(Aih)/ No when stopped taking drug(Damdawi lak zawh rualin a tawp)

d. If nor, When did it stop? (Engktikah nge a tawp?)

_____

e. Any medicine taken to reduce the side effect (with name of the medicine? (Side effect ti ziaawm turin damdawi ei/lak I nei em? Damdawi hming?)

_____

II. Gastric Cancer Associated Adverse Drug Reaction(Side effect lanchhuah dan):

ADR	**CTCAE GRADE**					**Remark**
	Grade 1 Mild	Grade 2 Moderate	Grade 3 Severe	Grade 4 Life Threatening	Grade 5 (Death)	
Adrenal insufficiency						
Anorexia						
Constipation						
Auditory (Loss of hearing)						
External Ear pain						

Internal Ear pain						
Otitis, external ear						
Change in appetite						
Colitis						
Dermatologic adverse reactions						
Atrophy						
Alopecia (Loss of Hair)						
Dry skin						
Rash						
Pruritus/itching						
Hyperthyroidism						
Endocrine disorder						
Diabetes						
Drowsiness						
Fatigue						
Gastrointestinal Tract Complication						
Constipation						
Dehydration						
Heartburn/dyspepsia						
Gastritis						
Mucositis - Anus – Esophagus – Large bowel – Larynx – Oral cavity – Pharynx – Rectum – Small bowel – Stomach – Trachea						
Necrosis, GI – Select: ---Necrosis, GI – Select – Anus – Colon/cecum/appendix – Duodenum – Esophagus – Gallbladder – Hepatic – Ileum – Jejunum – Oral – Pancreas – Peritoneal cavity – Pharynx – Rectum – Small bowel NOS						

– Stoma – Stomach						
Obstruction, GI – Select: – Cecum – Colon – Duodenum – Esophagus – Gallbladder – Ileum – Jejunum – Rectum – Small bowel NOS – Stoma – Stomach						
Stricture/stenosis (including anastomotic), GI – Select: Stricture, GI – Select – Anus – Biliary tree – Cecum – Colon – Duodenum – Esophagus – Ileum – Jejunum – Pancreas/pancreatic duct – Pharynx – Rectum – Small bowel NOS – Stoma – Stomach						
Hepatitis						
Hypersensitivity						
Hypophysitis						
Iron Overload in blood						
Platelets						
Splenic function						
Hypotension						
Pulmonary hypertension						
Restrictive cardiomyopathy						
Cardiac Arrhythmia						
Myelosuppression						
Nausea						
Nephritis						
Neurologic Problems						
Oral Ulceration						
Pneumonitis						
Vomiting						

Chyle or lymph leakage						
Edema: head and neck						
Edema: limb						
Edema: trunk/genital						
Edema: viscera						
Acidosis						
hypoalbuminemia						
Alkaline phosphatase						
Alkalosis						
serum glutamic pyruvic transaminase						
hyperbilirubinemia						
hypocalcemia						
Arthritis						
Exostosis						
Joint-effusion						
Memory impairment						
Mood alteration – Agitation – Anxiety – Depression – Euphoria						
Neuropathy: sensory						
Phrenic nerve dysfunction						
Vision dysfunction						
Allergy						
Other specific Observation						

III. Diagnosed with any other disease after surgery? ?(In zai hnuah natna dangin a tlak buak che em?)

Disease	Phenotype	Medications	Still Continuing	Not Continuing (for how many days medication was prescribed)	Still persisting	Cured

IV) Family History of Adverse Drug Reaction (ADR Observed? Yes No

(I chhungte ah damdawi huat bik nei an awm em): ☐ ☐

Name:

Relation: ……. ☐ **FDR** ☐ **SDR**

Gender: M ☐ **F** ☐

Age when ADR observed: ☐

ADR Observed:

Other Specific Observations as well as if any of your family members have adverse reaction/ allergy to certain drugs (i chhungte ah damdawi huat bik nei an awm em):

(Food Drug Interaction Study)

7. Diet (General Information):

a) Chicken (Local/Broiler)?
Once in a week Two/three days per week All days in a week

b) Beef?
Once in a week Two/three days per week All days in a week

c) Pork?
Once in a week Two/three days per week All days in a week

d) Vegetables?
Once in a week Two/three days per week All days in a week

e) Fruits?
Once in a week Two/three days per week All days in a week

f) Fish?
Once in a week Two/three days per week All days in a week

g) Egg?
Once in a week Two/three days per week All days in a week

8. Other life Style habits:

i. Habit of Smoking/Drinking/Chewing pan-masala:

ii.

Type		**Daily**	**Once in a week** (How many Pegs/Sticks)	**Two to three times in a week** (How many Pegs/Sticks)
Smoking				
Alcohol				
Gutkha				
Pan-masala				
Betel-Nut (kuhva)				
Tobacco				
Tuibur				

iii. Sleep Habit:

a. Wake-up time (Morning)? (Zing dar engzatah nge I thawh?) ----------

b. **Do you sleep at day time? (Chhunah I mu ngai em?)** Yes No (----------Hours in case Yes)

c. Sleep Time at Night? (Zanah dar engzatah nge I mut?) ----------

d. Do you have a habit of eating anything right before Sleep? ?(I mut hmain thil ei leh I ching em?)
----

e. Duration from Dinner to sleep time? (Zanriah to mut hun/darkar engzah nge?) ----------

Aganezov S, Yan SM., Soto DC, Kirshe M, *et al*. (2022). A complete reference genome improves analysis of human genetic variation.Science; 376:eabl3533.

Altemose N, Logsdon G, Bzikadze A., Sidwani P, *et al.* (2022). Complete genomic and epigenetic maps of human centromeres.Science; 376:eabl4178.

Andrews. S. (2010). FastQC: a quality control tool for high throughput sequence data. Available online at: http://www.bioinformatics.babraham.ac.uk/projects/fastqc; Accessed 22 September 2022.

Asadov C, Aliyeva G, Mustafayeva K. (2017). Thiopurine S-Methyltransferase as a Pharmacogenetic Biomarker: Significance of Testing and Review of Major Methods. Cardiovasc Hematol Agents Med Chem., 15(1):23-30.

Assi S, Torrington E, Cheema E, Hamid AA. (2021). Adverse drug reactions associated with chemotherapeutic agents used in breast cancer: Analysis of patients' online forums. Journal of Oncology Pharmacy Practice., 27(1):108-118.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR, 1000 Genomes Project Consortium. (2015). A global reference for human genetic variation. Nature, 526(7571):68-74.

Auton A, Brooks LD, Durbin RM, Garrison EP, Kang HM, Korbel JO, Marchini JL, McCarthy S, McVean GA, Abecasis GR. (2015). 1000 Genomes Project Consortium, A global reference for human genetic variation. Nature., 526(7571):68-74.

Bennett CL, Hoque S, Olivieri N, Taylor MA, Aboulafia D, Lubaczewski C, Bennett AC, Vemula J, et al. (2020). Consequences to patients, clinicians, and manufacturers when very serious adverse drug reactions are identified

(1997-2019): A qualitative analysis from the Southern Network on Adverse Reactions (SONAR). E Clinical Medicine., 31:100693.

Bolger AM, Lohse M, Usadel B. (2014). Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics., 30(15):2114-20.

Božina N, Bilić I, Ganoci L, Šimičević L, Pleština S, Lešnjaković L, Trkulja V. (2022). *DPYD* polymorphisms c.496A>G, c.2194G>A and c.85T>C and risk of severe adverse drug reactions in patients treated with fluoropyrimidine-based protocols. Br J Clin Pharmacol., 88(5):2190-2202.

Brown, TA. (2002). Chapter 1: The Human Genome. Genomes (2nd edition). Oxford: Wiley-Liss.. Available from: https://www.ncbi.nlm.nih.gov/books/NBK21134/

Bush WS, Crosslin DR, Owusu-Obeng A, et al. (2016). Genetic variation among 82 pharmacogenes: The PGRNseq data from the eMERGE network. *Clin Pharmacol Ther*., 100(2):160-169.

Cerami E, Gao J, Dogrusoz U, Gross BE, Sumer SO, Aksoy BA, Jacobsen A, Byrne CJ, et al. (2012). The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discov., 2(5):401-4.

Chopra D, Rehan HS, Sharma V, Mishra R. (2016). Chemotherapy-induced adverse drug reactions in oncology patients: A prospective observational survey. Indian J Med Paediatr Oncol., 37(1):42-46.

Chorley BN, Wang X, Campbell MR, Pittman GS, Noureddine MA, Bell DA. (2008). Discovery and verification of functional single nucleotide polymorphisms in regulatory genomic regions: Current and developing technologies. Mutation Research/Reviews in Mutation Research, 659:(1-2), 147-157.

Coleman JJ, Pontefract SK. (2016). Adverse drug reactions. Clin Med (Lond)., 16(5):481-485.

Conyers R, Devaraja S, Elliott D. (2018). Systematic review of pharmacogenomics and adverse drug reactions in paediatric oncology patients. Pediatr Blood Cancer.,65(4): e26937.

Crawford DC, Nickerson DA. (2005). Definition and clinical importance of haplotypes. Annu Rev Med., 56:303-20.

Daneshi N, Holliday E, Hancock S, Schneider JJ, Scott RJ, Attia J, Milward EA. (2017). Prevalence of clinically actionable genotypes and medication exposure of older adults in the community. Pharmgenomics Pers Med.,10:17-27.

Dean L. (2012). Azathioprine Therapy and TPMT and NUDT15 Genotype. [Updated 2020 Aug 5]. Edited by Pratt VM, Scott SA, Pirmohamed M, et al., Medical Genetics Summaries. Available from: https://www.ncbi.nlm.nih.gov/books/NBK100661/.

Dere WH, Suto TS. (2009). The role of pharmacogenetics and pharmacogenomics in improving translational medicine. *Clinical Cases in Mineral and Bone Metabolism*, *6*(1), 13-16.

Duan F, Song C, Wang P, Ye H, Dai L, Zhang J, & Wang K. (2021). Polygenic Risk Scores for Prediction of Gastric Cancer Based on Bioinformatics Screening and Validation of Functional lncRNA SNPs. Clin Transl Gastroenterol., 12(11):e00430.

Dyhl-Polk A, Vaage-Nilsen M, Schou M, Vistisen KK, Lund CM, Kümler T, Appel JM, Nielsen DL. (2020). Incidence and risk markers of 5-Fluorouracil and capecitabine cardiotoxicity in patients with colorectal cancer. Acta Oncol., 59(4):475-483.

Edwards IR, Aronson JK. (2000). Adverse drug reactions: definitions, diagnosis, and management. Lancet., 356(9237):1255-9.

Etienne-Grimaldi MC, Boyer JC, Beroud C, Mbatchi L, van Kuilenburg A, Bobin-Dubigeon C, et al. (2017). New advances in DPYD genotype and risk of severe toxicity under capecitabine. PLoS One., 12(5):e0175998.

Finkelstein J, Friedman C, Hripcsak G, Cabrera M. (2016). Pharmacogenetic polymorphism as an independent risk factor for frequent hospitalizations in older adults with polypharmacy: A pilot study. Pharmgenomics Pers Med., 9, 107-116.

Gagan J, Van Allen EM. (2015). Next-generation sequencing to guide cancer therapy. *Genome Med*., 7, 80.

Gallant JE, Koenig E, Andrade-Villanueva J, Chetchotisakd P, DeJesus E, et al. (2013). Cobicistat versus ritonavir as a pharmacoenhancer of atazanavir plus emtricitabine/tenofovir disoproxil fumarate in treatment-naive HIV type 1-infected patients: week 48 results. J Infect Dis., 208(1):32-9.

Gammal RS, Court MH, Haidar CE, Iwuchukwu OF, Gaur AH, Alvarellos M, Guillemette C, Lennox JL, et al. (2016). Clinical Pharmacogenetics Implementation Consortium (CPIC) Guideline for UGT1A1 and Atazanavir Prescribing. Clin Pharmacol Ther., 99(4):363-9.

Gammal RS, Pirmohamed M, Somogyi AA, Morris SA, Formea CM, Elchynski AL, Oshikoya KA, McLeod HL, Haidar CE, et al. (2023). Expanded Clinical Pharmacogenetics Implementation Consortium Guideline for Medication Use in the Context of G6PD Genotype. Clin Pharmacol Ther. 113(5):973-985.

Gao J, Aksoy BA, Dogrusoz U, Dresdner G, Gross B, Sumer SO, et al. (2013). Integrative analysis of complex cancer genomics and clinical profiles using the cBioPortal. Sci. Signal. 6(269):pl1.

Gayden T, Mirabal S, Cadenas, A. et al. (2009). Genetic insights into the origins of Tibeto-Burman populations in the Himalayas. J Hum Genet 54, 216–223.

Gentile G, Botticelli A, Lionetto L, Mazzuca F, Simmaco M, Marchetti P, Borro M. (2016). Genotype-phenotype correlations in 5-Fluorouracil metabolism: a candidate DPYD haplotype to improve toxicity prediction. Pharmacogenomics J., 16(4):320-5.

Ghosh D, Skinner MA, Laing WA. (2007). Pharmacogenomics and nutrigenomics: synergies and differences. *Eur J Clin Nutr*., 61(5):567-574.

Gianni M, Dentali F, Lonn E. (2009). 5 flourouracil-induced apical ballooning syndrome: a case report. Blood Coagul Fibrinolysis.,20(4):306-8.

Giuliani C, Sazzini M, Pirazzini C, Bacalini MG, Marasco E, Ruscone GAG, Fang F, et al. (2018). Impact of demography and population dynamics on the genetic architecture of human longevity. Aging (Albany NY).,10(8):1947-1963.

Goekkurt E, Al-Batran SE, Mogck U, Pauligk C, Hartmann JT, Kramer M, Jaeger E, Ehninger G, Stoehlmacher J. (2009). Pharmacogenetic analyses of hematotoxicity in advanced gastric cancer patients receiving biweekly fluorouracil, leucovorin, oxaliplatin and docetaxel (FLOT): a translational study of the Arbeitsgemeinschaft Internistische Onkologie (AIO). Ann Oncol., 20(3):481-5.

Guo Y, Dai Y, Yu H, Zhao S, Samuels DC, Shyr Y.(2017). Improvements and impacts of GRCh38 human reference on high throughput sequencing data analysis. Genomics; 109(2):83-90.

Haas, BJ, Dobin, A, Li, B. *et al.* (2019). Accuracy assessment of fusion transcript detection via read-mapping and *de novo* fusion transcript assembly-based methods. *Genome Biol*; 20, 213.

Hacker M. (2009). Chapter 13- Adverse Drug Reactions, Pharmacology:Principles and Practice, Edited by Miles Hacker, William Messer and Kenneth Bachmann, Academic Press, pp. 327-352.

Hamid AB., Petreaca RC. (2020). Secondary Resistant Mutations to Small Molecule Inhibitors in Cancer Cells. Cancers (Basel).,12(4):927.

Hamzic S, Schärer D, Offer SM, Meulendijks D, Nakas C, Diasio RB, Fontana S, Wehrli M, Schürch S, Amstutz U, Largiadèr CR. (2021). Haplotype structure defines effects of common DPYD variants c.85T > C (rs1801265) and c.496A > G (rs2297595) on dihydropyrimidine dehydrogenase activity:

Implication for 5-Fluorouracil toxicity. Br J Clin Pharmacol., 87(8):3234-3243.

Hari A, Zhou Q, Gonzaludo N, Harting J, Scott SA, Qin X, Scherer S, Sahinalp SC, Numanagić I. (2023). An efficient genotyper and star-allele caller for pharmacogenomics. Genome Res.,33(1):61-70.

Hindorf U, Lindqvist M, Peterson C, Söderkvist P, Ström M, Hjortswang H, Pousette A, Almer S. (2006). Pharmacogenetics during standardised initiation of thiopurine treatment in inflammatory bowel disease. Gut. 55(10):1423-1431.

Houlahan KE, Livingstone J, Fox NS, Kurganovs N, Zhu H, Sietsma Penington J, Jung C, et al. (2023). A polygenic two-hit hypothesis for prostate cancer. J Natl Cancer Inst., 115(4):468-472.

Ingelman-Sundberg M, Rodriguez-Antona C. (2005). Pharmacogenetics of drug-metabolizing enzymes: implications for a safer and more effective drug therapy. Philos Trans R Soc Lond B Biol Sci., 360(1460):1563-1570.

Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, et al. (2021). IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes. *Nucleic Acids Res*., 49(D1):D1225-D1232

Johnson JA. (2001). Drug target pharmacogenomics: an overview. Am J Pharmacogenomics., 1(4):271-81.

Kabir R, Vasquez N, Keshvani N, Vongpatanasin W. (2020). A case of chemotherapy-induced coronary vasospasm in a patient with colorectal cancer. J Cardiol Cases.,22(3):117-120.

Kadoyama K, Miki I, Tamura T, Brown JB, Sakaeda T, Okuno Y. (2012). Adverse event profiles of 5-Fluorouracil and capecitabine: data mining of the public version of the FDA Adverse Event Reporting System, AERS, and reproducibility of clinical observations. Int J Med Sci., 9(1):33-9.

Karki R, Pandya D, Elston RC, Ferlini C. (2015). Defining "mutation" and "polymorphism" in the era of personal genomics. BMC Med Genomics., 8:37.

Karthikeyan K, Sameera KV, Shaji S, Swetha MAC, Madhu CS. (2022). Capecitabine induced Steven-Johnson syndrome: A rare case report. J Oncol Pharm Pract., 28(1):250-254.

Katara P, Yadav A. (2019). Pharmacogenes (PGx-genes): Current understanding and future directions. *Gene*.,718:144050.

Katara P. (2014). Single nucleotide polymorphism and its dynamics for pharmacogenomics. Interdiscip Sci., 6(2):85-92.

Kim W, Cho YA, Kim DC, Lee KE. (2022). Elevated Risk of Fluoropyrimidine-Associated Toxicity in European Patients with *DPYD* Genetic Polymorphism: A Systematic Review and Meta-Analysis. J Pers Med., 12(2):225.

King RJ, Yu F, Singh PK. (2017). Genomic alterations in mucins across cancers. Oncotarget., 8(40):67152-67168.

Klein TE, Ritchie MD. (2018). PharmCAT: A Pharmacogenomics Clinical Annotation Tool. Clin Pharmacol Ther.,104(1):19-22.

Kown SJ. (2005). Management of side effects of 5FU based chemotherapy. Korean J Clin Oncol.,1(1): 51-58.

Krebs K, Milani L. (2019). Translating pharmacogenomics into clinical decisions: do not let the perfect be the enemy of the good. *Hum Genomics*., 13(1):39.

Król ZJ, Dobosz P, Ślubowska A, Mroczek M. (2023). WGS Data Collections: How Do Genomic Databases Transform Medicine? Int J Mol Sci., 24(3):3031.

Kwok, AJ, Mentzer A, Knight JC., (2.21). Host genetics and infectious disease: new tools, insights and translational opportunities. *Nat Rev Genet.,* 22(3),137–153.

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

Landrum MJ, Kattman BL.(2018). ClinVar at five years: Delivering on the promise. Hum Mutat, 39(11):1623-1630.

Laura E. Russell, Yitian Zhou, Ahmed A. Almousa, Jasleen K. Sodhi, Chukwunonso K. Nwabufo & Volker M. Lauschke. (2021). Pharmacogenomics in the era of next generation sequencing – from byte to bedside. Drug Metab Rev., 53(2): 253-278.

Lesko LJ, Woodcock J. (2004). Translation of pharmacogenomics and pharmacogenetics: a regulatory perspective. Nat Rev Drug Discov., 3(9):763-9.

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

Li H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. https://doi.org/10.48550/arXiv.1303.3997.

Liao WW, Asri M, Ebler J, *et al.* (2023). A draft human pangenome reference. *Nature* 617, 312–324.

Lindor NM, Thibodeau SN, Burke W. (2017). Whole-Genome Sequencing in Healthy People. *Mayo Clin Proc*., 92(1):159-172.

Lonsdale, J., Thomas, J., Salvatore, M. et al. (2013) The Genotype-Tissue Expression (GTEx) project. Nat Genet., 45, 580–585.

Lopes Cardozo JMN, Andrulis IL, Bojesen SE, Dörk T, Eccles DM, Fasching PA, Hooning MJ, Keeman R, et al. (2023). Associations of a Breast Cancer Polygenic Risk Score With Tumor Characteristics and Survival. J Clin Oncol., 41(10):1849-1863.

Lu ZK, Samuel J Kessler BA, Schulz R, Bian J, Brian Chen JD, et al. (2014) Systematic Approach to Pharmacovigilance beyond the Limits: The Southern Network on Adverse Reactions (SONAR) Projects. Adv Pharmaco epidemiol Drug Saf, 3: 149.

Luzzatto L, Ally M, Notaro R. (2023). Glucose-6-phosphate dehydrogenase deficiency. Blood., 10;136(11):1225-1240.

Ma C, Wei S, Song Y. (2011). T790M and acquired resistance of EGFR TKI: A literature review of clinical reports. J Thorac Dis, 3(1), 10-18.

Ma MK, Woo MH, McLeod HL. (2002). Genetic basis of drug metabolism. Am J Health Syst Pharm, 59(21), 2061-2069.

Makałowski W. (2001). The human genome structure and organization. Acta Biochim Pol.;48(3):587-98.

Manolio TA, Collins FS, Cox NJ, et al. (2009). Finding the missing heritability of complex diseases. *Nature*.461(7265):747-753.

Marieke J.H. Coenen, Aimée D.C. Paulussen, Marc Breuer, Martijn Lindhout, Demis C.J. et al. (2019). Evolution of Dihydropyrimidine Dehydrogenase Diagnostic Testing in a Single Center during an 8-Year Period of Time,Current Therapeutic Research, 90:1-7.

Martínez-Jiménez F, Muiños, Sentís I. et al. (2020). A compendium of mutational cancer driver genes. Nat Rev Cancer., 20, 555–572.

Maruyama T, Fuerst PA. (1985). Population bottlenecks and nonequilibrium models in population genetics. II. Number of alleles in a small population that was formed by a recent bottleneck. Genetics.,111(3):675-89.

Matáková T, Halašová E, Škovierová H, Dzian A, Dobrota D, Škereňová M. (2017). *DPYD* genotype and haplotype analysis and colorectal cancer susceptibility in a case-control study from Slovakia. Gen Physiol Biophys., 36(5):557-563.

McKenna A, Hanna M, Banks E, Sivachenko A, Cibulskis K, Kernytsky A, et al. (2010). The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data. *Genome Res.*, 20(9):1297-303.

Meyerson W, Leisman J, Navarro FC, *et al*. (2020). Origins and characterization of variants shared between databases of somatic and germline human mutations. BMC Bioinformatics 21(1), 227.

Molden E, Jukić MM. (2021). CYP2D6 Reduced Function Variants and Genotype/Phenotype Translations of CYP2D6 Intermediate Metabolizers: Implications for Personalized Drug Dosing in Psychiatry. Front Pharmacol., 12:650750.

Moraes F, Góes A. (2016) A decade of human genome project conclusion: Scientific diffusion about our genome knowledge. *Biochem Mol Biol Educ*.;44(3):215-223. doi:10.1002/bmb.20952.

Morales-González E, Fernández J, Pong-Wong R, Toro MÁ, Villanueva B. (2021). Changes in Allele Frequencies When Different Genomic Coancestry Matrices Are Used for Maintaining Genetic Diversity. Genes (Basel).,12(5):673.

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

NIH/National Human Genome Research Institute. Researchers Trim Count of Human Genes To 20,000-25,000. ScienceDaily. Retrieved March 2, 2023. Available from www.sciencedaily.com/releases/2004/10/041021075155.htm.

Nunez-Torres R, Pita G, Peña-Chilet M, López-López D, Zamora J, Roldán G, Herráez B, Álvarez N, Alonso MR, Dopazo J, et al. (2023). A Comprehensive Analysis of 21 Actionable Pharmacogenes in the Spanish

Population: From Genetic Characterisation to Clinical Impact. Pharmaceutics., 15(4):1286.

Nurk S, Koren S, Rhie A, *et al.* (2022), The complete sequence of a human genome. Science, 376, 44-53.

Offer SM, Lee AM, Mattison LK, Fossum C, Wegner NJ, Diasio RB. (2013). A *DPYD* variant (Y186C) in individuals of african ancestry is associated with reduced DPD enzyme activity. Clin Pharmacol Ther. 94(1):158-66.

Ostroverkhova D, Przytycka TM, Panchenko AR. (2023). Cancer driver mutations: predictions and reality. Trends Mol Med., 17:S1471-4914(23)00067-9.

Pavlos R, Mallal S, Phillips E. (2012). HLA and pharmacogenetics of drug hypersensitivity. Pharmacogenomics., 13(11):1285-306.

Pellicer M, García-González X, García MI, Robles L, Grávalos C, García-Alfonso P, Pachón V, Longo F, et al. (2017). Identification of new SNPs associated with severe toxicity to capecitabine. Pharmacol Res., 120:133-137.

Preissner SC, Hoffmann MF, Preissner R, Dunkel M, Gewiess A, Preissner S. (2013). Polymorphic Cytochrome P450 Enzymes (CYPs) and Their Role in Personalized Therapy. PLoS ONE, 8(12):e82562.

Puerta-García E, Urbano-Pérez D, Carrasco-Campos MI, Pérez-Ramírez C, Segura-Pérez A, Calleja-Hernández, Cañadas-Garre M. (2020). Effect of DPYD, MTHFR, ABCB1, XRCC1, ERCC1 and GSTP1 on chemotherapy related toxicity in colorectal carcinoma. Surg Oncol., 35:388-398.

Puerta-García E, Urbano-Pérez D, Carrasco-Campos MI, Pérez-Ramírez C, et al. (2020). Effect of *DPYD, MTHFR, ABCB1, XRCC1, ERCC1* and *GSTP1* on chemotherapy related toxicity in colorectal carcinoma. Surg Oncol., 35:388-398.

Qing T, Mohsen H, Marczyk M, Ye Y, Zhao H, Townsend J P, et al. (2020). Germline variant burden in cancer genes correlates with age at diagnosis and somatic mutation burden. Nature Communications, 11(1), 1-8.

Relling MV, Klein TE, Gammal RS, Whirl-Carrillo M, Hoffman JM, Caudle KE. (2020). The Clinical Pharmacogenetics Implementation Consortium: 10 Years Later. Clin Pharmacol Ther.,107(1):171-175.

Roden DM, McLeod HL, Relling MV, et al. (2019). Pharmacogenomics. *Lancet*., 394(10197):521-532.

Roden DM, McLeod HL, Relling MV, Williams MS, Mensah GA, Peterson JF, Van Driest SL. (2019). Pharmacogenomics. *The Lancet*, *394*(10197), 521-532.

Ruzzo A, Graziano F, Galli F, Galli F, Rulli E, Lonardi S, Ronzoni M, Massidda B, Zagonel V, Pella N, Mucciarini C, et al. (2017). Dihydropyrimidine dehydrogenase pharmacogenetics for predicting fluoropyrimidine-related toxicity in the randomised, phase III adjuvant TOSCA trial in high-risk colon cancer patients. Br J Cancer., 117(9):1269-1277.

Samer CF, Lorenzini KI, Rollason V, Daali Y, Desmeules JA. (2019). Applications of CYP450 testing in the clinical setting. *Mol Diagn Ther*., 17(3):165-84.

Samuel GN, Farsides B. (2017). The UK's 100,000 Genomes Project: manifesting policymakers' expectations. New Genet Soc., 36(4):336-353.

Sangkuhl K, Whirl-Carrillo M, Whaley RM, Woon M, Lavertu A, Altman RB, Carter L, Verma A, Ritchie MD, Klein TE. (2020). Pharmacogenomics Clinical Annotation Tool (PharmCAT). Clin Pharmacol Ther., 107(1):203-210.

Sharma A, Kumari KM, Manohar HD, Bairy KL, Thomas J. (2015). Pattern of adverse drug reactions due to cancer chemotherapy in a tertiary care hospital in South India. Perspect Clin Res., 6(2):109-15.

Sharma R, Kellarai A. (2014). Pharmacovigilance and adverse drug reaction reporting perspectives among interns and postgraduates of a teaching hospital. J Pharmacol Pharmacother., 5(4):248-50.

Shenfield GM. (2004). Genetic Polymorphisms, Drug Metabolism and Drug Concentrations. Clin Biochem Rev., 25(4), 203-206.

Shenfield GM. (2004). Genetic Polymorphisms, Drug Metabolism and Drug Concentrations.nClin Biochem Rev., 25(4), 203-206.

Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, Sirotkin K. (2001) dbSNP: the NCBI database of genetic variation. Nucleic Acids Res, 29(1):308-11.

Shifu Chen, Yanqing Zhou, Yaru Chen, Jia Gu. (2019). fastp: an ultra-fast all-in-one FASTQ preprocessor, Bioinformatics., 34(17):i884–i890.

Shrestha S, Shakya R, Shrestha S and Shakya S. (2017). Adverse Drug Reaction due to Cancer Chemotherapy and its Financial Burden in Different Hospitals of Nepal.Int J Pharmacovigil., 2(1):1-7.

Singh M, Prasad CP, Singh TD, Kumar L. (2018). Cancer research in India: Challenges & opportunities. *Indian J Med Res*.;148(4):362-365.

Skrzypczak-Zielinska M, Borun P, Milanowska K, Jakubowska-Burek L, Zakerska O, et al. (2013). High-resolution melting analysis of the TPMT gene: a study in the Polish population. Genet Test Mol Biomarkers.,17(2):153-159.

Sotoudeh M, Shirvani SI, Merat S, Ahmadbeigi N, Naderi M, (2019). *MSLN* (Mesothelin), *ANTXR1* (*TEM8*), and *MUC3A* are the potent antigenic targets for CAR T cell therapy of gastric adenocarcinoma. J. Cell Biochem. 120(4):5010-5017.

Sotoudeh M, Shirvani SI, Merat S, Ahmadbeigi N, Naderi M. (2019). *MSLN* (Mesothelin), *ANTXR1* (*TEM8*), and *MUC3A* are the potent antigenic targets for CART cell therapy of gastric adenocarcinoma. J. Cell. Biochem., 120 (4), 5010–5017.

Stacy AE, Jansson PJ, Richardson DR. (2013). Molecular pharmacology of ABCG2 and its role in chemoresistance. Mol Pharmacol. 84(5):655-669.

Su W, Feng B, Hu L, Guo X, Yu M. (2022). *MUC3A* promotes the progression of colorectal cancer through the PI3K/Akt/mTOR pathway. BMC Cancer., 22(1):602.

Su W, Feng B, Hu L. et al. (2022). MUC3A promotes the progression of colorectal cancer through the PI3K/Akt/mTOR pathway. BMC Cancer 22, 602 (2022).

Swen JJ, Nijenhuis M, de Boer A, Grandia L, Maitland-van der Zee AH, Mulder H, et al. (2011). Pharmacogenetics: from bench to byte--an update of guidelines. Clin Pharmacol Ther., 89(5):662-73.

Tahsin Hassan Rahit KM, Tarailo-Graovac M. (2020). Genetic Modifiers and Rare Mendelian Disease. Genes, 11(3):239.

Tate JG, Bamford S, Jubb HC, Sondka Z, Beare DM, Bindal N, Boutselakis H, Cole CG, et al. (2019). COSMIC: the Catalogue Of Somatic Mutations In Cancer. Nucleic Acids Res., 8;47(D1):D941-D947.

Teh LK, Hamzah S, Hashim H, Bannur Z, Zakaria ZA, Hasbullani Z, Shia JK, Fijeraid H, et al. (2013). Potential of dihydropyrimidine dehydrogenase genotypes in personalizing 5-Fluorouracil therapy among colorectal cancer patients. Ther Drug Monit., 35(5):624-630.

Teh LK, Hamzah S, Hashim H, Bannur Z, Zakaria ZA, Hasbullani Z, Shia JK, et al. (2013). Potential of dihydropyrimidine dehydrogenase genotypes in personalizing 5-Fluorouracil therapy among colorectal cancer patients. Ther Drug Monit.,35(5):624-30.

Tirona RG, Kim RB. (2017). Introduction to Clinical Pharmacology. Clinical and Translational Science (Second Edition), Academic Press, Edited by: David Robertson D and Williams GH, 365-388.

Toffoli G, Giodini L, Buonadonna A, Berretta M, De Paoli A, Scalone S, Miolo G, Mini E, et al. (2015). Clinical validity of a DPYD-based pharmacogenetic test to predict severe toxicity to fluoropyrimidines. Int J Cancer.,137(12):2971-80.

Tornio A, Backman JT. (2018). Cytochrome P450 in Pharmacogenetics: An Update. Adv Pharmacol., 83, 3-32.

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.

Wahlang JB, Laishram PD, Brahma DK, Sarkar C, Lahon J, Nongkynrih BS. (2017). Adverse drug reactions due to cancer chemotherapy in a tertiary care teaching hospital. Ther Adv Drug Saf., 8(2):61-66.

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

Wang Q, Shashikant CS, Jensen, M. et al. (2017). Novel metrics to measure coverage in whole exome sequencing datasets reveal local and global non-uniformity. *Sci Rep* 7(1):885.

Wang T, Antonacci-Fulton L, Howe K, Lawson HA, Lucas JK, Phillippy HP. et al. (2022). The Human Pangenome Project: A global resource to map genomic diversity. Nature., 604(7906), 437.

Wang Z, Jensen MA, Zenklusen JC. (2016). A Practical Guide to The Cancer Genome Atlas (TCGA). Methods Mol Biol., 1418:111-41.

Witherspoon DJ, Wooding S, Rogers AR, Marchani EE, Watkins WS, Batzer MA, Jorde LB. (2007). Genetic similarities within and between human populations. Genetics., 176(1):351-9.

Yi H, Li H, Liang L, Wu Y, Zhang L, Qiu W, Jiang W, Yang F, Li Q, Yang Z, Wang C, Cui L, He Y. (2019). The glucose-6-phosphate dehydrogenase Mahidol variant protects against uncomplicated Plasmodium vivax infection and reduces disease severity in a Kachin population from northeast Myanmar. Infect Genet Evol., 75:103980.

Zanger UM, Klein K. (2013).Pharmacogenetics of cytochrome P450 2B6 (CYP2B6): advances on polymorphisms, mechanisms, and clinical relevance. Front Genet. 4:24.

Zhang F, & Lupski JR. (2015). Non-coding genetic variants in human disease. Human Mol Genet., 24(R1), R102.

Zhang H, Li YM, Zhang H, Jin X. (2007). DPYD*5 gene mutation contributes to the reduced DPYD enzyme activity and chemotherapeutic toxicity of 5FU:

results from genotyping study on 75 gastric carcinoma and colon carcinoma patients. Med Oncol., 24(2):251-258.

Zhao M, Ma J, Li M, Zhang Y, Jiang B, Zhao X, et al. (2021). Cytochrome P450 Enzymes and Drug Metabolism in Humans. Int J Mol Sci., 22(23).

Zhou SF, Liu JP, Chowbay B. (2009) Polymorphism of human cytochrome P450 enzymes and its clinical impact, Drug Metab Rev., 41:2, 89-295.

BIO-DATA

Name	**:**	RANJAN JYOTI SARMA, M.Sc, M.Phil
Gender	**:**	Male
Current Designation	**:**	Project Assistant.
Total Work Experience	**:**	5 Years 8 month NGS Data Analysis, Working Experience in Linux, Basic Programming Skills in Python.
Organization	**:**	Department of Biotechnology Mizoram University
Current Address	**:**	Tanhril, Aizawl, Mizoram, India, PIN: 796004
Phone	**:**	(+91) 97744 40639
Email	**:**	ranjanjsarma@gmail.com cbiorjs@gmail.com

About: Experienced Data Analyst with a focus on genomics, proficient in NGS and Linux, beginner in ML seeking opportunities to further develop skills in this field. I hold M.Sc. Bioinformatics and M.Phil. in Biotechnology (Transcriptomics), currently pursuing Ph.D. (Pharmacogenomics) published 6 articles and developed 4 tools with strong analytical and problem-solving abilities.

Current Work Details:

Year	Experience	Designation	Project	Organization
November 2022-Present	11 Months	Project Assistant	DBT-Advanced Level State Biotech Hub	Department of Biotechnology, Mizoram University

Past Work Details:

Duration	Experience	Designation	Project	Organization
November 2022 -	11 Months	Project Associate	DBT-Advanced Level State Biotech Hub	Department of Biotechnology, Mizoram University
February 2020 - October 2022	2 Year 8 Months	PA-Data Analyst	DBT-GenomeIndia: Cataloguing the Genetic Variation in Indians	
June 2019 -February 2020	8 Months	Junior Research Fellow	DBT-Bioinformatics Infrastructure Facility	
June 2017 - June 2019	2 Years	Junior Research Fellow	DBT-Advanced Level State Biotech Hub	

Educational Background:

Qualification	Board/University	Year	Division	%	CGPA
M.Phil. Biotechnology	Mizoram University, Mizoram, India	2019	-		8.5
M.Sc. Bioinformatics	Pondicherry University, Pondicherry, India	2017	I	75	7.5
B.Sc. Biotechnology	North Eastern Hill University, Shillong, India	2015	I	63	-
Class XII	Nalbari College (Board: AHSEC), Assam, India	2011	I	66	-
Class X	Bangaon H.S. School (Board: SEBA), Assam, India	2009	I	75	-

Tools/Software Developed:

1. **WEAP:** Automated Whole Exome Sequencing Data Analysis Pipeline: https://github.com/ranjanjs34/weap

2. **Fastqc2pdf:** Automated Generate PDF reports of FASTQC from large scale NGS raw data: https://github.com/ranjanjs34/fastqc_pdf

3. **QTQ:** Automatically Perform QC, Quality Trimming and re-run QC on large scale NGS raw data. **https:**//github.com/ranjanjs34/QTQ

4. **Variant_Extractor**: Automatically extract variants of list of genes from large scale annotated data and arrange it sample-wise. https://github.com/ranjanjs34/Variant_Extractor

Publications:

1. **Sarma, RJ**., et al., (2023). Novel germline variants of MUC3A in a patient with ER+ breast cancer and signet-ring cell stomach adenocarcinoma. Gene Reports, 33, 101803. https://doi.org/10.1016/j.genrep.2023.101803.

2. Mohan G, **Sarma RJ**, et al., (2021) **Highlighting novel genes associated with the classical Rett syndrome patient from India.** ***Genes Disease***,9(6):1394-1396. https://doi.org/10.1016/j.gendis.2021.12.002.

3. **Sarma RJ**, e al., (2022). **Transcriptome analysis reveals SALL4 as a prognostic key gene in gastric adenocarcinoma**. ***Journal of the Egyptian National Cancer Institute***, 34(1):11. https://doi.org/10.1186/s43046-022-00108-5.

4. Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, Senthivel V, Divakar MK, Rophina M, Jolly B, Batra A, Sharma S, Siwach S, Jadhao AG, Palande NV, Jha GN, Ashrafi N, Mishra PK, A K V, Jain S, Dash D, Kumar NS, Vanlallawma A**, Sarma RJ**, *et al.* (2021). **Genetic epidemiology of autoinflammatory disease variants in Indian population from 1029 whole genomes.** ***Journal of Genetic Engineering and Biotechnology***, 19(1):183. https://doi.org/10.1186/s43141-021-00268-2.

5. Sahana S, Sivadas A, Mangla M, Jain A, Bhoyar RC, Pandhare K, Mishra A, Sharma D, Imran M, Senthivel V, Divakar MK, Rophina M, Jolly B, Batra A, Sharma S, Siwach S, Jadhao AG, Palande NV, Jha GN, Ashrafi N, Mishra PK, Vidhya AK, Jain S, Dash D, Kumar NS, Vanlallawma A, **Sarma RJ** *et al.* (2021). **Pharmacogenomic landscape of COVID-19 therapies from Indian population genomes.** ***Pharmacogenomics***, 22(10):603-618. https://doi.org/10.2217/pgs-2021-0028.

6. Jain A., Bhoyar R.C., Pandhare K., Mishra A., Sharma D., Imran M., Senthivel V., Divakar M.K., Rophina M., Jolly B., Batra A., Sharma S., Siwach S., Jadhao AG., Palande NV., Jha GN., Ashrafi N, Mishra PK, Vidhya A. K., Jain

S., Dash D., Kumar NS, Vanlallawma A., **Sarma RJ**, *et al*., (2020). **IndiGenomes: a comprehensive resource of genetic variants from over 1000 Indian genomes.** ***Nucleic Acids Research***, 49(D1): D1225–D1232. https://doi.org/10.1093/nar/gkaa923.

7. Sarathbabu S, **Sarma RJ**, Lalhruaitluanga H, Velmurugan D, Subramanian S., Kumar N. S. (2019). **In vitro DNA binding activity and molecular docking reveals pierisin-5 as an anti-proliferative agent against gastric cancer**. ***Journal of Molecular Structure & Dynamics***; 38(12): 474–3488. https://doi.org/10.1080/07391102.2019.1659856.

Conference Presentation:

1. **Ranjan Jyoti Sarma**, Jeremy Lalrinsanga Pautu, Bawitlung Zothankima, Saia Chenkual ,John Zohmingthanga, Nachimuthu Senthil Kumar.(2022). Study of Genetic Variation in Important Pharmacogenes in Gastric Cancer Patients from Mizoram, India. APOCP11 Conference on Cancer Prevention of the Asian Pacific Organization for Cancer Prevention (APOCP) at Kolkata, India.

2. ***[Virtual]*** **Ranjan Jyoti Sarma,** Saia Chenkual, John Zohmingthanga ,Nachimuthu Senthil Kumar. (2022). Screening of 5-Fluorouracil toxicity associated genetic variants in Gastric Cancer patients from North-East India: A Pharmacogenomics approach. National Conference on Cancer Biology and Therapeutics 2022 (CBT 2022), organized by Department of Zoology, Patharkandi College, Karimganj, Assam and Department of Zoology, PDUAM, Eraligool, Karimganj, Assam, India.

3. **Ranjan Jyoti Sarma**, Nachimuthu Senthil Kumar. (2018). Differential gene expression profiling of Gastric cancer RNA-Seq Data. 12th Annual Convention of Association of Pharmacy and Biotechnology (ABAP) & International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Treands (BEHIET 2018) at Mizoram University, Mizoram, India.

Workshop/Seminars Attended:

1. **Biomedical Data Analysis using SPSS (2023).** DST Sponsored 7 Days Hands-on workshop on statistics conducted by Technology Enabling Centre, Mizoram University.

2. **Next generation Sequence (Exome) Data Analysis in Human Diseases**. DST Sponsored 7 Days hands-on workshop conducted by Department of Biotechnology, Mizoram University.

3. **International Conference on Data Science in Biology (ICDSB) 2020.** September 3, 4, 5 & 8, 2020.Jointly conducted by Indian Institute of Technology, Jodhpur and Institute of Bioinformatics, Bangaluru.

4. **R for Statistics and Data Science. 25th -27th August 2020.** Conducted by Department of Mathematics, Bishop Cotton Women's College, Bengaluru

5. **2nd UK- India Cancer Bioinformatics Workshop.** 31st -2nd November. Conducted by King's College London and ACTREC at ACTREC, Navi Mumbai.

6. **Analysis of Genome Scale Data from Bulk and Single-cell Sequencing**. 19th - 23rd Nov 2018, Conducted by EMBL-EBI & NIBMG at NIBMG, Kalyani.

7. **3rd Advanced Research Training Workshop on Understanding Human Disease and Improving Human Health Using Genomics-Driven Approaches**. 23rd - 31st July 2018, Conducted by NIBMG, Kalyani.

8. **The Concept and Application of Genomics in Clinical Medicine**. 11th August 2018, Conducted by CSIR- Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi and Dept. of Biotechnology, Mizoram University.

9. **Hands-on Training Workshop on Cancer Genomics**. 19th -23rd March 2018, Organized by DBT- NER Biotechnology/Bioinformatics Centre, Advanced Centre for Treatment, Research & Education in Cancer, Kharghar, Navi Mumbai.

10. **Recent Advances in Cancer Research-2018.** 5th -7th March 2018, organized by Department of Biosciences and Bioengineering, Indian Institute of Technology Guwahati and sponsored by Department of Biotechnology, Government of India.

11. **A Brief Introduction to Bioinformatics and Systems Biology**. 13th -14th December 2018, organized by Bioinformatics Infrastructure Facility (BIF), department of Biotechnology, Mizoram University sponsored by Department of Biotechnology (DBT), New Delhi.

12. **Statistical Methods in Biological Research**. 3rd – 5th November 2017, organized by Bioinformatics Infrastructure Facility (BIF), Department of

Biotechnology, Mizoram University sponsored by Department of Biotechnology (DBT), New Delhi.

13. **Application of NGS in Microbial Ecology**. 30th – 2nd November 2017, organized by Bioinformatics Infrastructure Facility (BIF), Department of Biotechnology, Mizoram University sponsored by Department of Biotechnology (DBT), New Delhi.

14. **Research Training Workshop on Understanding Human Disease and Improving Human Health Using Genomics-Driven Approaches.** 19th – 24th November 2017, Conducted by NIBMG AT NIBMG, Kalyani.

Skills Earned:

- NGS pipeline Development, Workflow Automation
- Programming Language: Bash, Python3, MySQL, R
- Transcriptomics, Pathway Analysis, Whole Exome Data Analysis, CNV Analysis, Pharmacogenomics Annotation.

Software/Tools Learned:

Data Analytics: SPSS, R Studio, MS Excel, LibreCalc.

NGS: Fastqc, Fastp, Trimmomatic, STAR, BWA, Tophat, Bowtie2, BBMap, Segemehl, Samtools, bcftools, vcftools, Control-FREEC, GATK HaplotypeCaller, Mutect2, Picard, Annovar, PharmCAT, Stargazer.

Permanent Address : Kalbari, Bangaon, Nalbari, Assam, India-781303

Language : English, Hindi, Assamese

Declaration:

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

Date **Faithfully**

29 October 2023 Ranjan Jyoti Sarma

PARTICULARS OF THE CANDIDATE

NAME OF THE CANDIDATE	:	Ranjan Jyoti Sarma
DEGREE	:	Ph.D.
DEPARTMENT	:	Biotechnology
TITLE OF THE THESIS	:	Genetic Variants in Pharmacogenes in North-East Indian population and their association with Adverse Drug Reaction in Gastric Cancer
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.	:	1703911
Ph.D. REGISTRATION NO. & DATE	:	MZU/Ph. D./1725 of 29.05.2020
EXTENSION (IF ANY)	:	N/A

Head

Department of Biotechnology

ABSTRACT

GENETIC VARIANTS IN PHARMACOGENES IN NORTH-EAST INDIAN POPULATION AND THEIR ASSOCIATION WITH ADVERSE DRUG REACTION IN GASTRIC CANCER

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

RANJAN JYOTI SARMA
MZU REGISTRATION NO.: 1703911
Ph. D. REGISTRATION NO.: MZU/Ph. D./1725 of 29.05.2020

DEPARTMENT OF BIOTECHNOLOGY
SCHOOL OF LIFE SCIENCES
OCTOBER 2023

GENETIC VARIANTS IN PHARMACOGENES IN NORTH-EAST INDIAN POPULATION AND THEIR ASSOCIATION WITH ADVERSE DRUG REACTION IN GASTRIC CANCER

BY

RANJAN JYOTI SARMA

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 haploid whole genome is more than 3 billion base-pair (bp) long with an interindividual similarity of approximately 99.5%. However, due to population admixture, individuals from different populations are likely to be more similar than those from the same population. The human reference genome provides centralized genomic coordinates that are useful for comparing genomic loci identified in other study results. Evolving technology in DNA sequencing played an unprecedented role in understanding genome diversity in terms of polymorphism in the genes in coding and non-coding regions.

Common diseases such as different types of cancer are influenced by more than one genetic mutation. There are several germline gene mutations identified using polygenic risk scores derived from genome-wide association studies in prostate cancer, gastric cancer, breast cancer, and other common forms of cancers. Genetic polymorphism not only causes the disease state complex but also affects the prognosis of the disease by various means. Disease prognosis also gets affected by the variations in the gene involved in the process of drug metabolism. Genes involved in drug metabolism are referred to as pharmacogenes or PGx-Genes and are also responsible for variable drug responses (VDR). VDR could be influenced by the common as well as rare polymorphisms across a population and could be studied for pharmacogenomics and pharmacogenetics. VDR is also one of the major causes of adverse drug reaction (ADR) due to any variable responses during the pharmacokinetic and pharmacodynamic pathways of the drug. Pharmacokinetics is defined by the body's overall actions on the drug which involves the absorption followed by distribution, metabolism, and excretion (ADME) of the drug molecule. SNPs in the CYP family genes may have impacted the pharmacokinetics of commercial drugs.

Polymorphism of CYP enzymes encoding genes has been a primary focus in pharmacogenetics since these enzymes are responsible for the metabolism wide range of marketed drugs. Particular attention in pharmacogenetic testing has been devoted to CYP2D6, CYP2C19, CYP2C9,

and CYP3A4/5 genes as they encode the most common and highly variable CYP enzymes involved in drug metabolism. It was found that less than 14% of Asians, Africans, and Caucasians also experienced CYP2D6 deficiency, due to polymorphisms, and are classified as poor metabolizers. CYP2C9 can be considered as the most clinically significant metabolizer as the multiple SNPs identified in the gene directly impact the efficacy of the drugs and are also responsible for ADR events. Moreover, several polymorphisms in CYP family genes are also involved in carcinogen bioactivation.

2. Adverse Drug Reaction (ADR)

Adverse Drug Reaction (ADR) was considered to be a global medical concern. In 2000, Edwards and colleagues defined ADR as a notably harmful or unpleasant response that might be arise from any medication. ADR indicates a potential risk if the medication is administered again and should be either prevented, treated, or addressed by changing the dosage or discontinuing the product. Similarly, it was defined by the World Health Organization (WHO) as "a response to a drug that is noxious, unintended, and which occurs at doses normally used in man for prophylaxis, diagnosis or therapy of disease or the modification of a physiological function" (Sharma et al., 2014). In clinical setups, ADRs are considered major healthcare problems. However, patients with mild reactions are likely able to complete a given treatment course in particular cases as mild ADRs are easily manageable.

3. ADRs in Cancer And the Role of SNPs

ADRs are common and unavoidable risks associated with cancer chemotherapy. Detecting, monitoring, and preventing ADRs are crucial aspects of cancer patient care. A prospective study in Kathmandu, Nepal, revealed that age over 60 and female gender were risk factors for ADR development due to anticancer medications. The primary ADR-causing drugs were alkylating agents and antimetabolites, with specific drugs like Carboplatin, Gemcitabine, and fluorouracil playing significant roles. Anaemia was the most prevalent ADR, and many ADRs persisted even after

discontinuing the suspected drug. Most ADRs were considered probable in causality, moderate in severity, and probably preventable. ADRs increase the cost of illness due to additional therapy, clinical investigations, and prolonged hospital stays. Managing ADRs remains a significant challenge in cancer patient care, necessitating vigilance, monitoring, and prevention to enhance pharmaceutical care for patients.

4. Problem Statement

The Asian population is distinctly diverse from the Western populations. Although pharmacogenomic testing is available for several drugs, there is a limitation that it might not be applicable in all populations. A considerable portion of medicines commonly prescribed are metabolized by enzymes coded by highly polymorphic genes. In Asia, North-east Indian populations showed a genetic affinity with Mongoloids from southeast Asia. North Indians and north-eastern populations are markedly unrelated. The Northeast Indian population exhibits significant genetic diversity compared to other regions. There could be the possible influence of genetic variations in drug responses in this population which is crucial for tailoring the treatment of the most prevalent diseases to individuals' needs. Despite understanding the importance of Pharmacogenes in invoking ADR, their polymorphisms are not studied in response to population level.

Furthermore, when examining cancer incidence data spanning from 2012 to 2016, sourced from the 11 Population-Based Cancer Registries (PBCRs), it becomes evident that the northeastern region of India carries the highest burden of cancer cases. Specifically, Aizawl and Kamrup Urban in Assam have consistently reported elevated cancer incidence rates since 2003, affecting both men and women alike. Of particular concern in this region is the prevalence of gastric cancer, which imposes a substantial healthcare challenge. Aizawl district in Mizoram stands out with the highest incidence of gastric cancer among men. Several risk factors contribute to this alarming trend, including Helicobacter pylori infection, advancing age, a diet high in salt, and the consumption of diets low in fruits and vegetables, as highlighted in to recent studies. Intriguingly, despite the pressing need for tailored treatment

approaches in the context of gastric cancer in the Northeast Indian population, there is a noticeable gap in comprehensive studies on pharmacogenetics. This knowledge gap prompted the initiation of the present study, which aims to rectify this deficiency by providing valuable data on pharmacogenetics, thereby contributing to the advancement of more effective gastric cancer treatment strategies in this region.

5. Objectives of the study

The objectives of the proposed study are:

1. Cataloguing the coding and non-coding Pharmacogene variants in the north-east Indian healthy population.
2. Identification of the clinically actionable Pharmacogene variants in the Mizo healthy population.
3. Adverse Drug Reactions (ADRs) in Gastric Cancer (GC) patients and identification of genetic variants.

6. Materials and Methods

In the current research, a group of 93 healthy volunteers from different states of North east India was carefully selected for a whole genome sequencing (WGS) and also 27 self-declared healthy volunteers from various regions of Mizoram, belonging to the Mizo tribe were recruited for whole exome analysis. These individuals were genetically unrelated, ensuring diversity in the study sample. Moreover, stomach cancer patients were identified after cancer diagnosis by oncologists and pathologists, they were approached with consent forms and structured questionnaires. The research was aimed to investigate genetic variations and to understand their probable affect on drug metabolism which may exert ADR in the patients undergoing chemotherapy for stomach cancer at the state. A two-fold study was conducted to achieve this: First, the study initially focused on assessing the genetic variation among the patients receiving chemotherapy for stomach cancer. It aimed to identify genetic factors that might influenced the response to widely prescribed chemotherapy. Secondly, a follow-up study was conducted specifically to monitor and

document ADRs experienced by gastric cancer patients who had undergone chemotherapy and grading was performed using the guidelines of Common Terminology Criteria for Adverse Events v3.0 (CTCAE). Out of the total number of patients who had received chemotherapy, 37 were successfully traced and followed up until 03 May 2023.

The samples collected for the whole genome were processed for genomic DNA isolation at CSIR-IGIB, New Delhi where genomic DNA was extracted using the salting out method and subsequent quality control measures were followed before sequencing. Similarly, the samples for WES were processed in an in-house laboratory for genomic DNA isolation. Genomic DNA had been taken from blood samples using the QIAamp® Blood Mini Kit (Lot. 51304, QIAGEN).

Whole genome sequence data pre-processing, alignment to reference genome and variant calling was performed using Illumina DRAGEN v3.4 Bio-IT platform (Illumina Inc. San Diego, CA, USA). The joint variant calling for 93 individuals was performed using the Sentieon pipeline that integrates GATK protocols for joint genotyping which calculates the, allele count, allele number and allele frequency for the variants.

The samples collected during the period from 2016 to 2019 were processed and sequenced at the National Institute of Biomedical Genomics located in Kalyani, India. These samples were sequenced in the NIBMG, Kalyani. Samples collected after the year 2019 were processed and sequenced at Neuberg Diagnostics, Ahmedabad, India.

WES data analysis was done using an HP Hi-End computing server, employing an internally developed automated pipeline known as WEAP, which stands for Whole Exome Analysis Pipeline. WEAP was specifically created for variant calling from trimmed FASTQ data and was designed to call both germline and somatic variants following the best practices outlined by GATK guidelines. It operates in two modes: serial mode, processing samples one at a time, and parallel mode, handling four samples simultaneously. Significantly, WEAP's parallel mode proved to be much faster when dealing with extensive sample sets. The pipeline WEAP incorporates various essential

tools, including BWA aligner, Samtools, Picard, GATK, bedtools, vcftools, and Annovar, thereby establishing it as a comprehensive and standard automated solution for variant calling from WES data.

VIPs, which are genes of paramount importance in pharmacogenomics, were sourced from PharmGKB (https://www.pharmgkb.org/vips, accessed: 1 May 2022. PharmGKB serves as a comprehensive repository encompassing genes, their various genetic variants, star alleles, phenotypic information, clinical guideline annotations, drug label annotations, clinical annotations, and variant annotations. Additionally, it includes data on relevant pathways and supporting literature. These essential genes, designated as VIPs, were meticulously extracted from PharmGKB and organized into a text file for further analysis and research purposes.

In order to provide a improved accessibility to gene variants within both whole exome and whole genome datasets, structured query language (SQL) was used. This strategy led to the development of a dedicated database known as MPVardb version 1. This database was carefully designed using MariaDB, Hypertecxt Preprocessor (PHP), HyperText Markup Language 5 (HTML5) with HyperText Markup Language (CSS), and was hosted on an Apache web server using XAMPP environment.

To annotate the pharmacogenomic variant, the current research utilized the joint genotyped with variant score recalibrated and filtered by GATK derived from healthy exome dataset. The data was annotated using PharmCAT (https://pharmcat.org/; Sangkuhl et al., 2020). PharmCAT generated the clinical report based on the pharmacogene genotypes that matched with the prescribing recommendations which can be used to inform treatment decisions.

The PK/PD pathway genes were subsequently also investigated for their impact on protein expression due to the genetic variant. eQTL allows to check how a genetic variant may impact the protein expression. The analysis was performed using the publicly available dataset hosted in the Genotype-Tissue Expression (GTEx) portal. GTEx-eQTL dashboard (https://gtexportal.org/home/eqtlDashboardPage) was used to run the analysis where the variant of the PK/PD of genes was fed in the form of GRCh38

coordinates.

A deeper exploration into the significant genes identified in the eQTL analysis to investigate their impact on the overall survival of patients with stomach cancer. This exploration was conducted using the KM-Plotter tool (https://kmplot.com), focusing on patients who had received specific chemotherapy as adjuvant treatment. To investigate the mutation frequencies of genes within the PK/PD pathway of chemotherapeutic drugs, we utilized publicly accessible datasets, including those from TCGA (The Cancer Genome Atlas), accessed through cBioPortal (https://cBioPortal.org). Finally, an assessment was done of overall survival, taking into account the mutation status of genes within the PK/PD pathway associated with chemotherapeutic drugs. We conducted this analysis using the TCGA STAD dataset, and the tool employed for this purpose was https://tcga-survival.com/.

7. Results and Discussion

The PharmGKB database has classified 67 genes as very important pharmacogenes (VIPs). These VIPs were further categorized into Tier 1 with 33 genes, Tier 2 with 25 genes, and Tier 3 with 9 genes. The number of VIP genes has recently increased to 68 genes with the addition of TYMS and is expected to grow more based on the role of the genes in clinically important therapeutics (Hewett et al., 2002).

The 67 genes were screened for the variants in healthy exome and NE Indigenomes datasets. In the exome datasets derived from 27 healthy volunteers, there were 490 variants within the coding, 1112 in intronic, 95 in 3'UTR, and 52 in 5'UTR regions. Among the variants, 163 variants were non-synonymous, and 315 variants were found to be synonymous. Moreover, 1 frame-shift substitution, 4 start-loss, 1 stop-loss, and 2 stop-gain (2) variants were found. About 77,799 variants were found in 67 VIPs in the NE IndiGenomes dataset. There were 2527 variants in the exonic region, whereas 67,144 variants were found in intronic regions. There were 2086 variants in 3'UTR3, 378 variants in 5'UTR, 107 variants in intergenic, 82 variants in ncRNA_exonic, 5445 variants in ncRNA_intronic, 24 variants in splicing, 5

variants in upstream, and 1 variant in exonic splicing region. Among the other variant types, 910 variants were synonymous, and 1462 variants were non-synonymous. Moreover, 59 frameshift, 40 non-frameshift, 8 start-loss, 3 stop-gain, stop-loss variants were also found. However, 8 variant types remained unknown. The visualization of the variant details is made available through Mizoram Pharmacogenomics Variant Database MPVardb v1.1 constructed using MariaDB v10.4.28 with XAMPP v 3.3.0.

A vast difference in the number of variants in exonic variants in both datasets was observed. The exonic variants in exome and NE Indigenomes datasets were found to be 490 and 2527 variants, respectively. Although exome represents only 2% of the genome and WES technology can sequence those coding regions, one of the major limitations of WES is the uneven coverage of the target regions by the sequence reads. Currently, MPVardb hosts all types of germline variants including novel variants and rare variants (MAF < 0.01) resulting from the WES and WGS experiments based on the Northeast Indian populations. The database provides the allele frequencies of the variants derived from the healthy exome from Mizoram (n=27) and from the NE Indigen datasets (n=93).

There were multiple important CA PGx variants were observed in the healthy exome datasets. A variant Chr4:88131171 G>T (rs2231142) in ABCG2 with allele frequency (AF) in the healthy exome and Indigen (NE datasets) projects were found to be 0.259 and 0.811, respectively. The variant is responsible for decreased function (DF) for the drugs allopurinol and rosuvastatin. There was no related star allele detected in the variant, however, PharmaGKB-DPWG's clinical recommendations were available.

Another variant Chr19:40991369 C>T (rs8192709) was detected CYP2B6 gene responsible for intermediate function (IF) for Efavirenz, Sertraline. The reference allele C is *1, while the altered allele T is *10 for the variant can also be represented as CYP2B6*10. The variant AF in healthy exome and IndiGen NE datasets were found to be 0.111 and 0.913, respectively. Another important gene TPMT variant Chr6:18130687 T>C (rs1142345) or TPMT*3A was detected in one donor (AF= 0.01). The variant

AF in the Indigen NE dataset was found to be 0.961. The variant is an intermediate metabolizer for thiopurine-based drugs Azathioprine, Mercaptopurine, and Thioguanine and both the PharmGKB-DPWG and CIPIC provided the recommendation with clinical guidelines.

Another variant ChrX:154534495 C>T in G6PD (rs137852314) or G6PD-Mahidol variant was detected in the healthy exome that is responsible for variable function in the metabolism of drugs like Aspirin, Chloramphenicol Chloroquinon, Norfloxacin, exfloxacin, Quinine and many more. The AF in healthy exome and NE IndiGen dataset was found in 0.05 and 0.989, respectively.

The variant Chr2:233760498 G>A (rs4148323) or UGT1A1*6 (the altered allele denoted by *6) variant with AF 0.13 was detected in the healthy exome, affects effect the metabolism of Atazanavir, Irinotecan. The variant AF in indigen NE datasets was found to be 0.779.

The variant Chr12:21176804 A>G (rs2306283) or SLCO1B1*14 potentially leads to a decrease in the metabolism of statins. The variants AF in healthy exome and IndiGen NE dataset were found to be 0.722 and 0.354, respectively. Clinical guidelines for the variant have been provided by CIPIC for the variant SLCO1B1*14 (Table 4.1).

Two DPYD variants were detected by PharmCAT with normal function annotation. The variant Chr4:88131171 G>T (rs1801159) or DPYD*5 with AF in healthy exome and NE Indigen datasets were 0.255 and 0.725, respectively. Another variant Chr19:40991369 C>T (rs17376848) in DPYD was observed with AF in healthy exome and IndiGen NE datasets as 0.111 and 0.924, respectively. No star allele assigned to the variant rs17376848 yet. There are no clinical guidelines provided by PharmGKB-DPWG and CIPIC.

Historical migration patterns can significantly impact allele frequencies. The Mizo population's unique migration history or prolonged isolation may have led to distinctive allele frequencies. Conversely, the NE IndiGen dataset could include populations with different migration histories, further contributing to the observed differences. Small population size is another factor influencing the disparities in star allele frequencies. In smaller

populations, genetic drift can play a substantial role in allele frequency fluctuations. Random events can lead to the loss or fixation of specific alleles, contributing to differences between the Mizo healthy exomes and NE IndiGen datasets.

Genetic drift and founder effects can further compound these disparities. Over time, in smaller populations, random fluctuations in allele frequencies can become more pronounced, potentially leading to the observed differences. Founder effects, which occur during the initial establishment of a population, can also amplify these disparities. To ascertain the precise factors driving these variations in star allele frequencies, additional genetic research, population genetics analyses, and historical investigations are warranted. Furthermore, conducting functional studies can help elucidate whether these differences have any biological significance in terms of health or adaptation within these distinct populations.

ADRs are the most important causes of morbidity and mortality and increase the economic burden on patients and society (Sharma et al., 2015). In this study, 37 gastric cancer patients were followed up and 29 ADRs of different grades were documented. The portion of males and females showing different ADRs were 75.68% and 24.32%, respectively. Loss of Appetite was found to be a priominent ADR which was observed in 26 patients, followed by the occurrence of Fatigue in 21 patients, Nausea in 23 patients, drowsiness in 18 patients, Dermatological Adverse Reaction in 11 patients, Vomiting in 8 patients, Alopecia in 8 patients, and many more. Neutropenia, as often can be seen as ADR in patients undergoing anti-cancer therapy was also observed in 3 patients. Similarly, Leukopenia in 2 patients and low hemoglobin count in 1 patient. However, these ADRs were not observed as singletone. The patient showing Loss of Appetite also showed other ADR as comorbidities such as nausea, and drowsiness with others.

It is essential for healthcare providers to closely monitor patients, manage side effects proactively, and adjust treatment when necessary to optimize the therapeutic benefits while minimizing discomfort and risks associated with these reactions. Effective communication between patients and

their healthcare team is vital in addressing and managing these challenges, ensuring the best possible outcomes in the fight against stomach cancer. Furthermore, pharmacogenomics research can identify the potential genetic variants that induce such toxicities in patients.

The study found that most gastric cancer patients (about 92%) at least 5FU and 72% of it recieves 5FU with leucovorin and oxaliplatin. This made it really important to check for certain gene variations that affect how the body handles 5-FU, as 5FU undergoes enzymatic breakdown to exert its effect. On the other hand, oxaliplatin undergoes non-enzymatic breakdown. These gene variations can impact how effective the treatment is and whether there might be any side effects. Therefore, genetic screenings can provide valuable insights into individualized therapy approaches, ensuring that patients receive the most suitable and tailored treatments while mitigating potential adverse drug reactions, thereby advancing the field of gastric cancer management.

Three specific genes, DPYD, DPYS, and UPB1, play essential roles in the way the body processes and responds to 5-Fluorouracil (5-FU). DPYD, for example, is responsible for breaking down 5-FU, and variations in this gene can affect how quickly or slowly the drug is metabolized, potentially leading to variations in its effectiveness and the risk of side effects. DPYS and UPB1 also contribute to the complex process of how 5-FU interacts with the body's biological pathways. Moreover, the TYMS gene plays an important role in the pharmacodynamics of 5-FU. Understanding these genetic factors is crucial in tailoring 5-FU treatments to individual patients, ensuring optimal therapeutic outcomes while minimizing potential adverse reactions.

The study found that most gastric cancer patients (about 92%) at least 5FU and 72% of it recieves 5FU with leucovorin and oxaliplatin. This made it really important to check for certain gene variations that affect how the body handles 5-FU, as 5FU undergoes enzymatic breakdown to exert its effect. On the other hand, oxaliplatin undergoes non-enzymatic breakdown. These gene variations can impact how effective the treatment is and whether there might be any side effects. Therefore, genetic screenings can provide valuable insights into individualized therapy approaches, ensuring that patients receive the most

suitable and tailored treatments while mitigating potential adverse drug reactions, thereby advancing the field of gastric cancer management.

Three specific genes, DPYD, DPYS, and UPB1, play essential roles in the way the body processes and responds to 5-Fluorouracil (5-FU). DPYD, for example, is responsible for breaking down 5-FU, and variations in this gene can affect how quickly or slowly the drug is metabolized, potentially leading to variations in its effectiveness and the risk of side effects. DPYS and UPB1 also contribute to the complex process of how 5-FU interacts with the body's biological pathways. Moreover, the TYMS gene plays an important role in the pharmacodynamics of 5-FU. Understanding these genetic factors is crucial in tailoring 5-FU treatments to individual patients, ensuring optimal therapeutic outcomes while minimizing potential adverse reactions.

The current study found the involvement of these three genes in 5-FU metabolism based on the literature. It has been cited that deficiency in any of these three genes may exert 5-FU-related adverse reactions (https://www.pharmgkb.org/; Hewett et al., 2002). These three genes were screened in 59 patients' annotated data for synonymous and nonsynonymous variants associated with deficiency of the respective enzymes. A total of 5 variants in DPYD and 2 variants in each in DPYS and UPB1 among 59 patients. Mutation status in the publicly available data in cBioPortal

5.4.1 Variants in DPYD and their Impact

Exome sequencing revealed that 12 patients (52.17%) out of 23 patients who received 5-Fluorouracil had DPYD variants responsible for DPD Deficiency. Among the DPYD variants, rs1801159 (DPYD*5), a nonsynonymous variant was found in 35% of the total number of patients. Among the patients who received fluorouracil with exome sequencing done (n=23), 7 patients were positive for rs1801159. Five of these patients (accounting for ~ 21.74% of the patients) were deceased and 2 patients were found alive. The variant is reported in ClinVar as being responsible for DPD deficiency and toxicity associated with 5-Fluorouracil. Four out of the five deceased patients received 5-fluorouracil with leucovorin and Oxaliplatin (6 to

8 Cycles) and one patient received oral capecitabine (4 Cycles).

The follow-up study was able to receive survival data from 3 deceased patients. One patient survived for less than a year, one patient survived for more than a year but less than two years and one patient survived for more than two years but less than three years. Six of these patients received intravenous 5-fluorouracil with leucovorin and oxaliplatin and one patient received oral Capicitaben. The patient who received capecitabine showed grade 3 constipation. Loss of appetite of varying degrees was observed in 5 patients. Fatigue of grade 2 was noted in 3 patients, grade 1 in one patient, and grade 2 fatigue in 3 patients. Additionally, two patients displayed grade 2 alopecia. Grade 2 nausea was documented in two patients, while one patient experienced grade 3 nausea, and another patient had grade 1 nausea. Additionally, one patient displayed mild diabetes, and another patient exhibited grade 3 thrombocytopenia, which is characterized by low blood platelet levels. Constipation was observed in only one patient. Mild to moderate drowsiness was observed in 3 patients. Similarly, mild to moderate vomiting was observed in two patients, and severe vomiting in one patient. No significant cis-eQTL was found for the variant rs1801159 which indicates that the variant may not have impact on the expression of the gene.

Although PharmGKB has depicted the variant of normal function, our study found the variant has a great impact on patients' health. Moreover, a study conducted by Zhang et al. (2007), polymorphisms of DPYD*5 (rs1801159) ws over represented in non-responsive of fluorouracil treated patient from Chinese population and suggested that DPYD*5 as probable predictors of the response to fluorouracil-based chemotherapy for gastric cancer patients. Another study found that 29.9% of the neutropenia cases were DPYD*5 positive (ANOVA, P-Value = 0.01). The study also suggested that DPYD*5 (rs1801159) potentially useful predictive markers of patients' responses to 5-FU chemotherapy (Teh et al., 2013). Therefore, a statistical test will be necessary in this particular population for their association with the adverse drug reaction.

A rare nonsynonymous DPYD variant rs112766203 was detected in one patient, which was also reported to be associated with DPD deficiency in

ClinVar. In the patient, there were various ADRs of different grades namely: Loss of Appetite (Grade 1), Nausea (Grade 1), Darkening of Nails (Grade 2), severe neutropenia, and Alopecia (Grade 1). The patient survived 512 days (17 months and 10 days) from the date of detection of cancer. The variant was predicted to be damaging by SIFT, Polyphen, and Mutation taster. The variant was also designated as probably damaging by Marieke et al. (2019). However, there is not enough evidence on the association of adverse events due to the variant in published literature.

Similarly, Another nonsynonymous variant of DPYD was rs1801160 (DPYD*6) observed in one patient. In a study conducted by Matáková et al. (2017), the variant was found to be associated with colorectal cancer. The variant was also found to be responsible for slowing down the degradation rate of 5-FU (Gentile et al., 2016). Another study revealed that the variant significantly induces fluorouracil-associated hematological toxicity in European patients (Kim et al., 2022). A recent study by Božina et al. (2022) suggested that the variant can be a potential candidate for the DPYD testing panel due its association with severe adverse drug reactions due to 5-FU treatment. The study also found two synonymous variants of DPYD, rs1801265, and rs17376848 in two patients. Although, the patients were found alive, both the variants were reported to be associated with DPD deficiency and the variants were also found to be associated with inducing ADRs in 5-FU-treated patients in other studies (Teh et al., 2013; Ruzzo et al., 2017; Puerta-García et al., 2020; Hamazic et al., 2021).

The eQTL analysis revealed that variants rs1801160 and rs1801265 also cause upregulation of DPYD and the patients who showed positive for the variants were also found alive. Gene expression-based survival analysis using GEO datasets in KM-Plotter revealed that upregulation of DPYD was associated with better overall survival of the patient (p-value=0.00047, HR=0.54). However, due the less allele frequency of these variants (less than 0.05), the variants may not be suitable for DPYD testing for the population.

Based on the mutation-based survival study based on TCGA data revealed poor survival outcomes with the patient with the DPYD mutant allele.

Moreover, looking at the AF of the variant rs1801159 in the 1000g, Mizoram healthy, NE IndiGenome as well as the occurrence of the variant in gastric cancer patients, the study suggests in-depth clinical research to generate evidence of association with ADRs and DPD enzyme activity which could serve as potential predictor for the response of 5-FU treatment that are being provided to the stomach cancer patients in the state of Mizoram.

Two synonymous variant of DPYS gene found in the followed up patients. The variant rs2298840 was detected in 11 out of 23 patients and the variant rs36087551 was detected in one patients along with rs2298840. Among the 11 patients, 7 patients (63%) weere found deceased. The variant rs2298840 was also found in 46% of patients out of all the 59 GC patients. The patient with the variant rs36027551 was also deceased and variant was not observed in the 1000g annotation datasets. The AF of the variants rs2298840 in 1000g datasets was 0.231629. The AF of the variants rs2298840 in 1000g datasets was 0.231629 (Table B). One patient (gc-35) had the two DPYS variants rs2298840 and NM_001385:exon3:c.C541T:p.R181W (rs36027551) along with one DPYD variant rs1801159 that are associated with DPD deficiency and Dihydropyrimidinase deficiency found to be diseased. It was observed that the majority of the deceased patients (n=6) survived less than 5 years (1825 days). Although some of the variants of DPYS were found associated with the 5-FU toxicity, there was no evidence of association the variants rs2298840 and rs36027551 with toxicity. Moreover, no significant eQTL was found for these variants. DPYS mutation status was found to be 2.2% in the stomach cancer data available in cBioPortal. The variant rs2298840 was present found in 11 patient of which 7 patient (~63%) patient did not survive. This finding led us to check the role of DPYS mutation in survival in TCGA data. It was observed that DPYS mutation might play role in poor outcome in gastric cancer patients. These findings necessiates further in-depth study on DPYS variants in stomach cancer patients treated with 5-FU.

Only one exonic variant (synonymous) NM_016327:exon10:c.G1122A:p.K374K (rs35916595) was observed in the UPB1 gene in the heterozygous condition in one patient (gc-79). The variant is reported for deficiency in bete-ureidopropionase, an enzyme that converts luoro-beta-ureidopropionate to fluoro-beta-alanine (FBAL). The AF of the variant in 1000g datasets was found to be 0.0042. Another variant in 5'-UTR NM_016327:c.-17A>T (rs2070475) was also in the same patient. The patient experienced grade 3 mucositis with neutropenia and survived for 548 days from the date of detection of the disease. Moreover, another 2 patients gc17 and gc-58 carrying the variant rs2070475 survived for 1665 days and 875 days, respectively. Different types of adverse reactions were observed in these patients including severe forms of anorexia, SOB, Diarrhoea and vomiting in gc17, mild forms of LoA, Anorexia, and diabetes in gc-58 and low haemoglobin (grade 2), mucositis (grade 3), and neutropenia (grade 3) in gc-79. Overall, 3 patients were found to be deceased out of 6 patients with UPB1 variants and 3 patients are still surviving. One of the alive patients (gc-83) experienced LoA (grade 3), fatigue (grade 2), nausea (grade 3), drowsiness (grade 3), DAR (grade 3), Alopecia (grade 3). The other alive patient experienced LoA (grade 2), Diarrhoea (grade 2), DAR (grade 3) and Alopecia (grade 2) ADRs. Interestingly, significant eQTL were found to be associated with the variant rs2070475. The variant caused upregulation of the UPB1 gene in many tissue types. To understand role of upregulation of UPB1 gene on survival, GEO datasets in KM-plotter were explored which suggested that the upregulation is not associtated with overall survival of the patients (P=0.057, HR=1.47). Furthermore, the survival data was not found in TCGA based on UPB1 mutation status. The mutation frequency of the gene is also less (0.8%) in the stomach cancer data available.

Furthermore, it was found that various ADRs were seen in the patient having more than one PK/PD gene variants. Grade 3 osteoporosis as ADR was observed in one patient (gc-2) and Loss of Appetite (grade 2), Fatigue (grade 2), and Nausea (grade 2) in another patient with variants rs1801159 in DPYD, rs2298840 in DPYS, and rs3786362 in TYMS. Loss of Appetite (grade 2),

Vomiting (grade 1), Dermatological Adverse Reaction (Grade 1), and Alopecia (grade 2) were observed in a patient (gc-35) with two synonymous variants in DPYS (rs36027551 and rs2298840) and nonsynonymous variants in DPYD (rs1801159). Constipation was observed in one patient with rs1801159 in DPYD and rs2298840 in DPYS. Similarly, two patients (gc 75 and gc76), rs3786362 in TYMS and rs1801159 in DPYD had experienced more than three ADRs. LoA (grade 3), Fatigue (grade 2), Nausea (grade 3), Vomiting (grade 2), Diabetes (grade 1), Drowsiness (grade 2), DAR (grade 3), Hypotension (grade 3) in one patient (gc-75), and LoA (grade 1), Fatigue (grade 1), Nausea (grade 1), Drowsiness (grade 1) were observed in another patient (gc-76).

Three variants rs13181 in ERCC2 (TG genotype), rs17376848 in DPYD (AG genotype), and rs1801133 in MTHFR (AG genotype) associated with FLO toxicity were found in 30.5%, 10% and 11.86% of GC patients, respectively. However, less toxic genotypes of the variants rs25487 in XRCC1 (CT genotype), rs1695 in GSTP1 (AG genotype), rs1799794 in XRCC3 (CT genotype), rs717620 in ABCC2 (CT genotype) and rs11615 in ERCC1 (AG genotype) were present in 88.13%, 27.11%, 1.69%, 38.98% and 98.30% of patients, respectively. The variants rs13181 in 2 patients and rs11615 in 3 patients occurred as singleton. The variants rs11615, rs25487 and rs717620 occurred as together in 16 patients (27%). Moreover, rs11615 and rs25487 occurred together in 13 patients (22%). The prevelance of the variants rs13181 (30.50%), rs25487 (88.13%), rs1695(27.11%), rs717620 (38.98%) and rs11615 (98.30%) were high among the 59 gastric cancer patients. However, all the five variants was not seen together in any patient. However, the variants rs1801133, rs25487 and rs11615 which are reviewed by the expert panel in ClinVar have occurred in 6 Patients (10%). Since these variants were already found to be associated with FLO toxicity in published literature and were also present in a group in the stomach cancer patient, further study is required to understand their role in developing toxicity.

8. **Scope of the Research**

- The study underscores the importance of considering genetic variants and their interactions in assessing the effectiveness and outcomes of adjuvant chemotherapy for gastric cancer patients.
- Understanding the genetic factors influencing treatment responses may lead to the development of more personalized and targeted therapies for gastric cancer patients in the future.
- The current finding of the study further suggest to unravel the complex genetic underpinnings of treatment responses in gastric cancer patients and potentially improve patient outcomes.
- The case study performed on the patient with ER+ breast cancer and Stomach Adenocarcinoma found statistically significant association of MUC3A variants with stomach cancer development within the Mizo patients and can put the population susceptible to stomach cancer

9. Limitations of the Study

- It's important to recognize that these studies have a limitation: there are not many clinical experts who are experienced in using pharmacogenetics.
- So, even though we tried to focus our survey on people interested in this field, fewer of them said they use pharmacogenetics at their organization.
- Moreover, due to the absence of electronic data capture medium for Adverse events, it was challenging to collect all kinds of real-time adverse events during the therapy.
- A substantial number of patients were untraceable which resulted in a smaller number of samples and due to which a statistical analysis could not be performed to identify potential genetic factors for ADR.