Genetic variation studies in candidate genes related to Type 2 Diabetes in Mizo population, Mizoram

By
FREDA LALROHLUI
Department of Biotechnology

Submitted

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

Department of Biotechnology
School of Life Sciences
MIZORAM UNIVERSITY
(A Central University Accredited with "A Grade" by NAAC)
Aizawl - 796 004, Mizoram, India

Dr. N. Senthil Kumar Ph.D
Professor

Mobile: +91-9436352574
Email: nskmzu@gmail.com

CERTIFICATE

This is to certify that the thesis entitled "**Genetic variation studies in candidate genes related to Type 2 Diabetes in Mizo population, Mizoram**" to Mizoram University for the award of the degree of Doctor of Philosophy Biotechnology by **Freda Lalrohlui** Registration No. **MZU/Ph.D/ 841 of 22.04.2016,** Ph.D scholar in the Department of Biotechnology, under my guidance and supervision and has not been previously submitted for the award of any degree in any Indian or foreign University. She has fulfilled all criteria prescribed by the UGC (Minimum Standard and Procedure governing Ph.D. Regulations). She has fulfilled the mandatory publication (Publication enclosed) and completed Ph.D. course work.

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

Place:MZU

Date:3rd October

(Prof. N. Senthil Kumar)
Supervisor

Declaration of the Candidate

I, Freda Lalrohlui, hereby declare that the subject matter of this thesis entitled **"Genetic variation studies in candidate genes related to Type 2 Diabetes in Mizo population, Mizoram"** is a 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 had not been submitted by me for any research degree in any other University/Institute. This is being submitted to Mizoram University for the award of the degree of Doctor of Philosophy in Biotechnology.

Date: (Freda Lalrohlui)

Candidate

(Dr. Th. Robert Singh)

Head

(Prof. N. Senthil Kumar)

Supervisor

Acknowledgements

- First and foremost, I would like to thank **GOD** for giving me strength, health and endurance to complete this work.
- I would like to acknowledge my indebtedness and render my deepest gratitude to my supervisor **Prof. N. Senthil Kumar,** Biotechnology, Mizoram University for his guidance and support and for being a great mentor throughout my research.
- I am also thankful to **Dr.Th. Robert Singh**, Head, Department of Biotechnology and **all the faculties** and **non- teaching staff** for their help, support and kind cooperation.
- I would also like to express my deepest gratitude to **all the participants** who volunteered for this study.
- I would also like to thank **Advanced Level State Biotech Hub** (BT/04/NE/2009) facility and **Bio-informatics Infrastructure Facility** (No.BT/BI/12/060/2012(NER-BIF-MUA), sponsored by Department of Biotechnology (DBT), New Delhi, Govt. of India for infrastructure support to carry out this work successfully .
- I am thankful to Department of Science and Technology, New Delhi for providing **DST-INSPIRE fellowship** (DST/INSPIRE/03/2015/001832).
- I owe my gratitude to **Dr. John Zohmingthanga**, Deputy Medical Superintendent, Civil Hospital, Aizawl and **Genesis Laboratory**, Dawrpui, Aizawl.
- To all my lab mates and research scholars at Department of Biotechnology for their support.
- Last but not the least, I would like to thank **my family** for their constant support and prayers and I would like to dedicate my work to my grandparents, **Rev Tlanglawma (L) and Thansiami (L).**

Dated:

Place: Aizawl, Mizoram **Freda Lalrohlui**

Abbreviations and Symbols

~	Approximately
°C	Degree celcius
≥	Greater than or equal to
≤	Less than or equal to
%	percentage
±	Plus or Minus
A	Adenine
ANOVA	Analysis Of Variance
BAM	Binary Alignment Map
BLAST	Basic Local Alignment Search Tool
BMI	Body Mass Index
bp	Base Pair
C	Cytosine
CI	Confidence Interval
DNA	Deoxyribonucleic Acid
dNTPs	deoxyribonucleotide triphosphate
e.g.	Exempli gratia : For example
et al.	*et alii*: and others
etc.	et cetera: and other things
G	Guanine
GOD-POD	Glucose Oxidase-peroxidase
HGNC	HUGO Gene Nomenclature Committee
mg/dl	Milli gram per deciliter
ml	milli litre
mM	Milli Molar
MODY	Maturity Onset Diabetes of the Young
mtDNA	Mitochondrial Deoxyribonucleic acid

n	Number
NCBI	National Center of Biotechnology Information
ng/µl	nanogram per microlitre
nm	Nanometre
OMIM	Online Mendelian Inheritance in Man
OR	Odds Ratio
PCR	Polymerase Chain Reaction
pH	Negative Ion Of Hydrogen Ion Concentration
pmol	Pico-Mol
rCRS	revised Cambridge Reference Sequence
RNA	Ribonucleic Acid
ROC	Receiver operating characteristic curve
SAM	Sequence Alignment Map
SD	Standard Deviation
SNP	Single Nucleotide Polymorphism
SPSS	Statistical Package for the Social Sciences
T	Thymine
T2DM	Type 2 diabetes Mellitus
TE	Tris Acetate EDTA

TABLE OF CONTENTS

List of Figures

List of Tables

Chapter I

Introduction

1. Introduction

Diabetes refers to a group of polygenic metabolic diseases which may be portrayed by hyperglycemia leading to the alterations in carbohydrate, protein as well as fat metabolism, which in turn leads to the defects in the action of insulin (ADA, 2012). Once the food is ingested, it gets digested in to smaller components like carbohydrates and starch which can be converted to maltose by the action of amylase enzyme, which may be further broken down in to monosaccharide such as glucose. Due to the metabolism of carbohydrate in our body, glucose molecules are formed and passed on to the blood stream (Galgani et al, 2015).

Insulin is a hormone that helps the cells to take up and utilizethe available glucose molecule from the blood (Banting et al., 1991). Insulin is secreted by the Pancreas, which is a leaf like organ located between stomach and responsible for both endocrine and exocrine function. Islets of Langerhans comprises of a cluster of cells present in pancreas which secrete "insulin". When the status of the blood sugar level increase after intake of food, the Insulin stored in the vacuoles are released in the blood by a process known as exocytose. Insulin takes part in the conversion of glycogen in the liver and promotes anabolism and storage of fats in the adipose tissue (Gillespie, 2006). The alteration in the production and mis-regulation of this insulin hormone and the response of the target cells may lead to excessive glucose in the blood stream which may finally lead to diabetes mellitus.

The history of diabetes dates back around 1550 BC where it was mentioned in an Egyptian papyrus about a rare disease which results in weight loss and frequent urination. In 250 BC, Apollonius of Memphis is credited to coin the term 'diabetes'

which refers to a disease where the amount of fluid discharge is more than which a patient may consume. Later on Matthew Dobson identifies that due to excess of sugar in the blood and urine, the people with diabetes have a sweet taste in the urine (Mac Cracken et al., 1997). In the early 1900s, the name "insulin", was proposed independently by Sir Edward Albert Sharpey-Schafer and Jean de Meyer. Diabetes was differentiated by Sir Harold Percival in 1936 as type 1 and type 2 based on the degree of sensitivity of insulin. Insulin is a natural hormone and one of the most important antibiotic medications in glycemic control level. It contributes certain part in diagnosis as reverse insulin resistance, improves lipid profile, reduce the lipotoxicity, anti-inflammatory effects and anti- platelets effects apart from its role in glycemic control (Galloway et al., 1994). Apart from insulin, certain drugs have been discovered- carbutamide is the first oral drug discovered in 1955 for the treatment of diabetes. Metformin which have a herbal lineage of *Galega officinalis* has become the most preferred oral agent in the administration of type 2 diabetes (Clifford et al, 2017). Certain research has been done on the drug through the ages, however the research was boosted when the use of the drug metformin was initiated in USA in 1995. In 1998, UK Prospective Diabetes Study (UKPDS) gives metformin a positive direction as the most preferred primary agent in the management of hyperglycemia in type 2 diabetes (McCreight et al., 2016).

The classification of diabetes mellitus consists of two broad groups, Type 1 and Type 2 diabetes mellitus. However, several other classifications have also been introduced based on their clinical and etiological characteristics (Kuzuya et al., 1997). The classification of diabetes among individual often depends on the present

situation of the individual during the time of diagnosis. There are also several situations where many diabetic individuals fail to fit in a single classification. This may be due to the change in hormones, steroids and certain other drugs where diabetes may be induced (ADA, 2013). Adult onset diabetes or type 2 diabetes mellitus is more common among two types of diabetes mellitus. Although the symptoms developed by individual with diabetes are typical, individuals with type 2 diabetes can go unnoticed. Early detection and treatment can prevent the risk of developing further complications. Some of the common symptoms listed by American Diabetes Association (ADA) in 2018 are:

- Frequent urination
- Increased thirst
- Increased Hunger- even though food is consumed
- Extreme tiredness
- Blurred vision
- Slow wound healing

The diagnoses as stated by World Health Organization (WHO) on type 2 diabetes are:

- 120 mg/dL or more of Fasting glucose level
- ≥200 mg/dL of Post prandial and random glucose level

HbA1c also known as Glycated hemoglobin level of 6.5% or more as a part of diagnostic criteria is still up for debate (ADA, 2010). There are many type of complications that arises when an individual is investigated and known to have type 2 diabetes. Though several complications in the skin due to bacteria, fungi, and itchiness can happen to anyone, an individual with diabetes is more prone to the

infection and is considered the first sign of the disease. Complications in the eye is common for people with the disease and may cause certain eye defects like glaucoma, retinopathy and cataracts which leads to a higher risk of blindness than individual without diabetes. Diabetic neuropathy occurs when the nerve is damaged due to long term duration of diabetes and can even lead to complications in the foot (Boulton et al. , 2005). Diabetic ketoacidosis (DKA) is a serious condition that occurs when ketone bodies accumulate in the body and eventually leads to diabetic coma or even death. When the sugar level is high in the blood, it wears out the kidney causing nephropathy and thus leading to excessive filtration of the blood. This wearing out of the kidney causes protein leakage through the urine called micro albuminuria (UAAS et al., 2017). The increase in blood pressure is also common among people with diabetes because when the blood pressure increases it puts the heart under pressure to pump more blood. This may cause a lot of stress to the heart leading to stroke, heart disease and other problems (Cho et al., 2015). HHNS also known as Hyperosmolar Hyperglycemic Non-ketotic Syndrome shows a serious medical condition most frequently seen in older individuals with uncontrolled diabetes. This syndrome arises when the sugar levels in the blood increases, stressing the body to eliminate the excess sugar in the form of urine (Pasquel et al., 2014). When the blood sugar level arises and uncontrolled, it effects and damage the work of the vagus nerve in the digestive tract by delaying either stopping or slowing down the removal of its contents leading to disorder known as Gastroparesis (Camilleri et al., 2013).

In 2014 it has been estimated that 422 million adults were living with diabetes, in which about 1.5 million deaths were reported in 2012 (Joshi et al., 2007). In 2017, over 72 million cases of diabetes were reported in India. Diabetes maybe a potential epidemic disease as more than 62 million individuals have been diagnosed with the disease. In the year 2000, with more than 32 million cases of diabetes mellitus, India topped the world with the highest number of incidence followed by China (20 million) (Kumar et al., 2013).

Type 2 diabetes develops due to the inability of the pancreas to produce adequate insulin or when resistivity towards insulin is attained by the body. Although different factors may take part in the development, the exact reason for the development is still unknown. For type 2 diabetes there is still no known cure (Abdulfatai et al., 2012). The risk of diabetes may likely developed due to obesity, family history of diabetes, lack of physical activity etc. The progression of the disease may be higher with individuals belonging to certain ethnicities; however the reasons are not fully understood. It may be associated with strong familial genetic predisposition (Zimmet et al.,1991; Harris et al., 1995; Valle et al., 1997). However, the genetics involved in this form of diabetes are not clearly defined and complex (Courten et al., 1997; Knowler et al., 1993). Although certain factors may help in contributing to the advancement of type 2 diabetes, Socio Economic Status (SES) may take part in the development of the disease as it involve better access to health services, occupational opportunities and individual life style and habits (Brown et al., 2004). According to several studies, the socio-economic status in developed countries varies inversely with the risk, prevalence and complications of type 2

diabetes. The nutritional inadequacies, psychological stress and unhealthy lifestyle of the individuals belonging to low economic status may lead to a substantial rise in the pervasiveness of type 2 diabetes due to the increase rate in physical inactivity and obesity (Feinstein et al., 1993; Malmstromet al., 1999). Several pathogenic processes may add contribution to the development of diabetes which may include processes in which the beta cells present in pancreas is destroyed which may result in insulin resistivity and deficiency. The involvement of *Staphylococcus aureus* is found to be associated with a common complication of diabetes like Diabetic foot ulcers (Bader et al., 2008). In some studies, they suggest that poor glycemic control may lead to the increase in risk for Tuberculosis (Leung et al., 2008; Pablos-Mendez et al., 1997) while other studies obliged (Leegaard et al., 2011).

In developing nations, the highest number of individuals diagnosed with diabetes falls in the age range between 45 to 65 years, while in developed nations the highest prevalence is found in those aged 65 years and above. When compared between men and women the rate of prevalence is similar, where the prevalence in men is slightly higher < 60 years of age as compared to women > age 65 years (Weill et al., 2004). The epidemiological studies published lack proper representation among different geographic and socioeconomic regions which leads to the over or under estimation of the common occurrence of the disease in India (Sadikot et al., 2004). Among the different ethnic groups in the world, diabetes prevalence can be observed and studied from a broad spectrum where the differences and unpredictability of different ethnic groups may be ascribe to cultural and environmental factors apart from genetics. Even though the difference in the

predominance of the disease may vary considerably among different ethnic population which may share similar environment, this assists the idea of the involvement of genetic factors leading to the predisposition of the disease. Aside from genes, individuals within a family share environments, habits and culture, still aggregation of the disease within the family is another source of evidence for genetic contribution to the disease (Swapan et al., 2006). Proteins are biochemical molecules which perform certain type of functional activities like immune responses, cell cycle, cell signaling and cell adhesion. They have mechanical or structural functions in the body and proteins consist of an enzyme which acts as biochemical catalyst (Berg et al., 2002). Considerable amount of research has been done with the possible alterations in glucose and lipid metabolism in type 2 diabetes patients. However, the studies involving protein metabolism may be sometimes conflicting (Staten et al., 1986; Luzi et al.,1993; Welle et al.,1990; Halvatsiotis et al.,2000).

Mutations in the genomic either, pathogenic or non-pathogenic, may also contribute to the development of the disease. Non-pathogenic mutations may include several single nucleotide polymorphisms though they have been studied and reported to act in association with type 2 diabetes (Haghvirdizadeh et al., 2015). Genome – wide association studies (GWAS), linkage studies, candidate gene approach etc. have identified several genes that have contribution to the susceptibility of Type 2 diabetes (Cornelis et al., 2009). The strategies for the treatment and prevention are needed urgently to put the disease at base (Lyssenko et al., 2008). Several linkage as well as candidate gene studies reported that many chromosomal regions are linked with type 2 diabetes and they have identified possible causative genetic variants in the

chromosome for the disease (Horikawa et al., 2000; Love-Gregory et al., 2004; Meyre et al., 2005). However, several variants which have been reported in certain populations cannot be reproduced in other populations; this may be attributed to the complexity of type 2 diabetes. Over the last few decades genome-wide association studies have been used to identify the common variants that may increase the risk of the common disease like type 2 diabetes. Hundreds of reproducible associations have been reported between common single SNPs and particular traits. Some of these associations have yielded novel biological insights that will be useful for biomedical research (Altshuler et al., 2008). In the human cells, the mitochondrial genome is the only cell that contains extra-chromosomal DNA. Any type of alteration in the mitochondrial genome has been previously acknowledged to be one of the causes of type 2 diabetes. Mutations which may be either rearrangement or point mutation were observed in families having history of type 2 diabetes (Ballinger et al., 1992; van den Ouweland et al., 1992).

Data on the prevalence of T2DM in Mizoram is limited and considering the great cultural, geographical and racial diversity of our country, there is a need for assessing the epidemiological and genetic factors that can potentially lead to diabetes in the state of Mizoram. To our knowledge, no studies have been reported focusing on the epidemiology, demography, and genetics from Mizo population. The present study on the Genetic variation studies in candidate genes related with Type 2 Diabetes in Mizo population, Mizoram has been proposed and designed which aims to study the demographic factors and clinico-biochemical profiles and to study mutations in candidate genes and their association with type 2 diabetes in Mizo Population. The findings in this study may help us to understand the influence of

these factors with type 2 diabetes, which may further help to a better understanding of the disease prevalence and may contribute to a better prognosis of the disease in this population.

Chapter II

Review of Literature

2. Review of Literature

The etiology of type 2 diabetes creates numerous debates over the centuries for researchers due to the complexity and nature of the disease. New discoveries are made in the diagnosis as well as prognosis of the disease. However, this new discoveries cannot be applied to every individual as type 2 diabetes involves genetic as well as environmental effects and is a combination of certain known and unknown factors. Certain criteria and risk factors have been laid down which can lead to the progression of the disease, however, they may not be applicable to an individual belonging to a specific group of population. Numerous researches have been carried out in different parts of the world with different people belonging to a wide distribution of population, race and groups. Studies on the prevalence of the disease in a particular population or groups create greater challenges to researchers in this field due to the multifactorial nature of the disease. This chapter highlights on different research carried out in type 2 diabetes from different parts of the world which comprises on study groups belonging to different population, involving molecular approach, biochemical, demography and epidemiology.

2.1. Research on the History of Type 2 diabetes and its development

The mechanism of glycogen synthesis in the liver, reported by Claude Bernard- a French physiologist in the 19th Century creates a pavement in diabetic research. The removal of pancreas from a dog to induce diabetes creates a milestone in the disease research and the experiment was performed by Oskar Minkowski and Joseph von Mering in 1889 (Karamanou et al., 2016). The duration and care taken for the disease is important in individuals having known type 2 diabetes as chronic complications are related to time period to which the patient is diagnosed. Regardless

of many debates by scientists and physicians, they come to terms where certain factors act independently and cause risk factors in the progression of the disease. The common risk of developing the disease may be modifiable and non-modifiable-modifiable risks are those where a certain change in the lifestyle habits of an individual can prevent the acquirement of the disease i.e., it can be prevented. However, non-modifiable risk factors are the factors that have already been acquired mostly genetically. The non-modifiable risk factors include having a familial history, pre-diabetes, gestational diabetes, ethnicity, old age and other metabolic symptoms. The modifiable risk factors include obesity, hypertension, physical inactivity etc. (Chen et al., 2011).

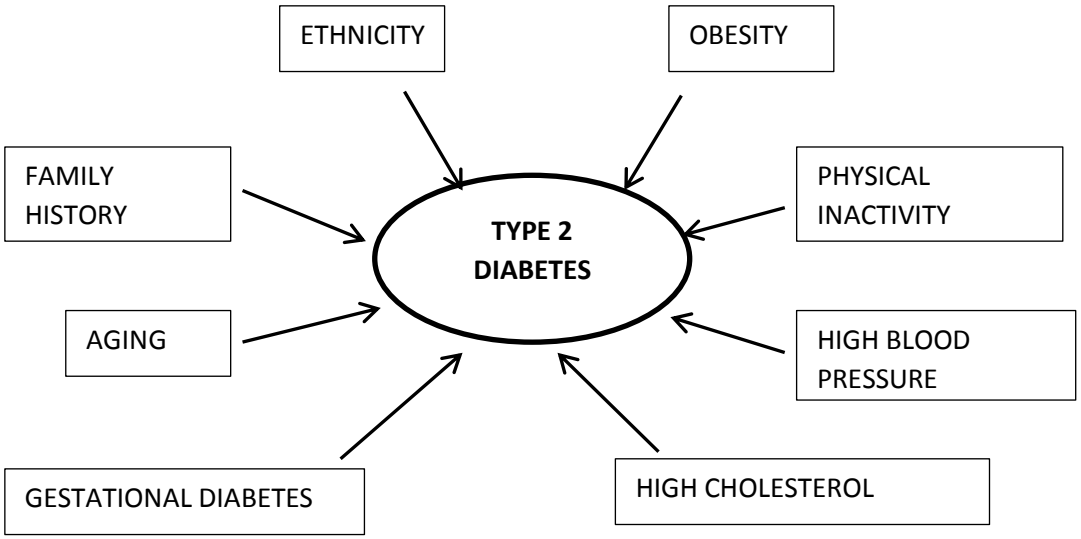


Figure 2.1: Common risk factors for type 2 diabetes development

The development of the disease may arise due to certain diabetogenic lifestyle such as obesity and physical inactivity with susceptible genotype. Pancreatic damage, certain medicine, other diseases and genetic mutation can lead to the progression as well as development of the disease. Monogenic diabetes occurs when a single gene undergoes mutation or changes and a number of these single gene

mutation causes alteration in insulin production in the pancreas. Majority of these changes is passed through families along the generation and some individual may developed this type of mutation without inheritance (Shields et al., 2010). Other diseases like cystic fibrosis and hemochromatosis affects the pancreas in such a way that it leads to the progression of type 2 diabetes. An autosomal recessive disorder called Cystic fibrosis can disfigure the pancreas as it produced thick mucus and defection in the conductance regulator of cystic fibrosis trans-membrane causes β-cell dysfunction (Kayani et al., 2018). Hemochromatosis is a condition where huge amount of iron is stored in the body and can eventually lead to pancreatic damage. This type of condition may also arise due to mutation in HFE genes and other associated genes (Capell et al., 2004) and activities like using oral contraceptives, excessive intake of vitamin C and iron, alcohol addiction and blood transfusions (Bovenschen et al.,2009). Certain hormonal diseases like Cushing's Syndrome, Acromegaly and Hyperthyroidism can cause insulin resistivity which in turn leads to type to diabetes. When the pancreas is damaged or destroyed, it can lead to several diseases like pancreatitis, trauma and pancreatic cancer. The defect in the pancreas causes the disruption of the beta cells to produce enough insulin for the body.

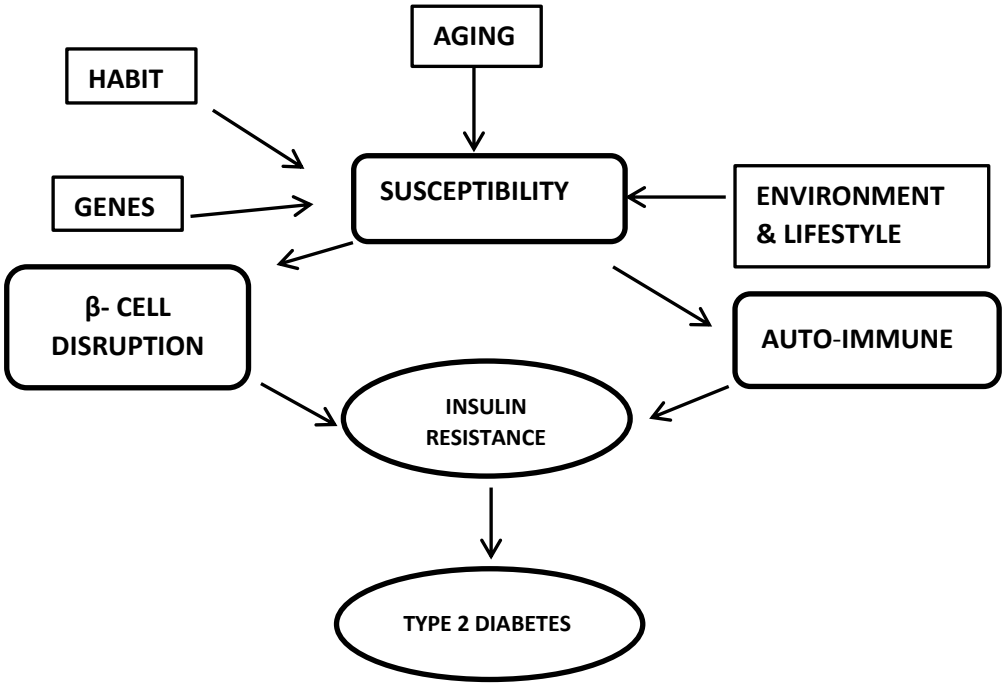


Fig 2.2: Schematic representation of type 2 diabetes development

Obesity is considered one of the major possibility factors for diabetes; however, the research related to this risk factor across India is yet to give meaningful results. Although India has lower rate of obesity and overweight, the prevalence of diabetes is higher in comparison to other countries which may suggests that even at much lower body mass index (BMI) diabetes may occur (Rao et al., 2011; Mohan et al., 2006).

2.2. Type 2 diabetes in relation to Environmental factors- Demography and epidemiology

There are several population based studies which focuses on the environmental factors leading to a particular disease in that population. Type 2 diabetes as reported from various research shows the influence of different environment

factors which can lead to the progression of the disease (Mariana et al., 2011; Meigs et al., 2000; Kaprio et al., 1992). In conjugation with some susceptible genes, some individuals with earlier behavioral factors like inactive lifestyles and diets rich in high fat content are likely to develop type 2 diabetes in their lifetime (Tuomilehto et al.,2001; Knowler et al.,2002). There are studies that shows that long time exposure to environmental pollutants either pesticides or herbicides can disturb the mechanism of glucose which further causes resistance towards insulin (Lee et al.,2006; Lee et al., 2007). Obesity may serve as a reservoir for toxic particles as the organic pollutants are lipophilic in nature and may gradually create insulin resistivity (Porta et al., 2006; Lee et al., 2006). However, studies shows that one risk alone may not be a causative agent for the prevalence of type 2 diabetes, but when it acts in association with other covariates like consistency of a population, educational status, ethnicity and smoking habits, the prevalence may increase (Brook et al.,2008; Pearson et al.,2010; Kramer et al.,2010). The risk of type 2 diabetes may also be attributed to the nature and status of the food consumed; whether they are free of chemical contaminants like dioxins, benzene, nitroso compounds, and polycyclic hydrocarbons as these compounds are found to be linked with the risk of acquiring the disease (Sun et al.,2009; Longnecker et al., 2001). In some studies, it has been hypothesized that antioxidants may be preventive against type 2 diabetes as oxidative stress plays a role in insulin resistivity (Ceriello et al., 2000). The biomarkers of dietary, fruits and vegetable consumption like carotenoids in the serum, vitamin C and tocopherol present in the plasma has opposite relationship with type 2 diabetes in relation to their concentration (Coyne et al., 2005, Salonen et al.,1995, Sargeant et al.,2000).

Among healthy individuals, it has been reported that there is a direct association between insulin sensitivity and the consumption of vitamin C and vitamin E (Facchini et al.,1996, Feskens et al.,1995).

The drastic change in our lifestyle and habits due to modernization leads to a great change in the development of the disease. It has been reported that the prevalence of type 2 diabetes cannot be justified to a particular age range as the disease also occurs at a relatively young age as well as old depending on the type of population and factors involved (Abdulghani et al.,2016). It has been reported that aging has taken part in the mechanism involved in sensitivity and resistivity of insulin known that aging (Meneilly and Elliot, 1999). It has been known from previous studies that type 2 diabetes may effect women excessively as they are known to have much lower glycemic control than men (Misra et al.,2009; Shalev et al.,2005; Tang et al.,2008; Chiu et al.,2011). However, the reason for prevalence of the disease in relation to sex still becomes a huge debate for researchers (Legato et al.,2006). There has been numerous questions regarding whether the difference in the body composition and hormones have something to do with the prevalence of type 2 diabetes between the male and female (Arnetz et al., 2014). One of the contributing risk factor from studies across the world reveals that the ethnicity in which an individual belongs can reveal the increase or decrease in the incidence of developing the disease. However, the contribution of ethnicity in the disease development may be attributed to the difference in environments as some ethnic groups or population has their own unique lifestyle and dietary practices (Abate et al., 2003).

From different epidemiological analysis, the risk factors that were discussed in developing type 2 diabetes like the type of food and quantity, socioeconomic

status, ethnicity etc., may contribute independently or at least take a major or partial part in the disease development (Kolb et al., 2017). Since type 2 diabetes develops due to the loss of insulin secretion, the environmental and different lifestyle factors may have a direct or indirect contribution to the destruction of β-cell (Butler et al., 2016).

2.3. Type 2 diabetes in relation to Protein, Amylase and Microalbumin

Since many protein play a key role in the body, there is a possibility that certain defect in the synthesis of some proteins may lead to insulin resistivity (Moller et al., 2008). It has been reported that elevation in the protein level can predict the development of type 2 diabetes (Pradhan et al., 2001). The salivary glands produced Ptyalin and the pancreas produce the pancreatic amylase an alpha-amylase which is present in the digestive systems of human. Amylase plays a role in the breaking down of complex molecules into simpler molecules (Taniguchi et al.,2009). Several findings shows that increase in the blood glucose level, duration of hyperglycaemia and older individuals with type 2 diabetes tends to have lower levels of amylase (Ishii et al.,2018; Yadav et al.,2013). The elevation in blood glucose levels can cause the fluctuation of amylase levels in the body. Since type 2 diabetes can cause damage to several organs in the body and pancreas being one of them, the production of pancreatic amylase by the pancreas may be altered (Aughsteen et al., 2005). Microalbuminuria is the presence of protein albumin in the urine and is considered one of the foremost marker for diabetic nephropathy (Rose et al.,2004). This may be due to the damage in kidney function that leads to the leakage of large amount of protein in the urine (Yuyan et al.,2004). Since the development of type 2 diabetes

alone is complex, the incidence of diabetic nephropathy becomes uncertain (Salah et al.,2002).

2.4. Type 2 diabetes in relation to genetic factors

Although there are many possible factors which play a role in the progression and evolution of type 2 diabetes, many individuals have more susceptibility in developing the disease. This is due to the fact that they inherited the susceptible genes from one of their parents or may be both. Before the 1980's, the genetic variants involved in the hereditary risk remains unclear but with the advancement in genetic research and technologies, several variants were discovered (Ali et al., 2013). From different studies among population, the chances of heritability of an individual to type 2 diabetes may range from 20 to 80 % ; 40% chance for a person with single parent diagnosed with the disease and 70% chance if both the parents are affected with the disease (Tillil et al.,1987; Meigs et al.,2000; Poulsen et al.,1999;). In comparison between positive and negative family history, individuals having positive family history especially with first degree relatives are three fold more likely to develop the disease (Florez et al.,2003). Studies where parents in the age range 35 to 60 years are focused, it has been observed that the risk of hereditary is higher. This also explain the role of environmental factors like lifestyle and dietary habits in the development of the disease at a later stage in life (Almgren et al.,2011). The genetic risk of type 1 diabetes may be concentrated on the HLA region of Chromosome 6, however in type 2 diabetes; the genetic risk is not concentrated to a particular region but rather the cooperation of a number of genes from different parts of the genome (Gibson et al.,2011). There are many questions and debate regarding the genetic component of type 2 diabetes as a number of common genetic variants may effect in

small proportions or maybe that the affect may be due to rare variants in small or large proportions.Although multiple genes are involved in the development of type 2 diabetes, there are those genes that have gained attention like potassium inwardly rectifying channel, subfamily J, member 11 (KCNJ11), and the peroxisome proliferator-activated receptor-gamma (PPARG) gene. The reason for the constant observation of these genes may be due to the fact that they are involved in the action and metabolism of insulin and glucose and other metabolic functions (Schwenk et al.,2013). Alteration in these genes such as mutation can lead to the disruption of KATP activity which can lead to a permanent diabetes mellitus from younger age (Abujbara et al.,2014). Adipogenesis and the development of insulin resistivity are found to be implicated with PPARG gene and deleterious and pathogenic mutations in this gene can cause the lack of response to insulin (Pattanayak et al.,2014). Maturity onset diabetes of the young (MODY) constitutes about less than 1% of all type 2 diabetes (Winckler et al.,2007). In the pancreas, HNF1A gene controls the transcription of HNF4A gene and mutation in HNF4A gene may lead to MODY1 while mutation in HNF1A may lead to MODY 3.This two MODY subtypes have similar clinical characteristics which may sometimes be difficult to distinguish (Boj et al.,2001). Dysfunction in the pancreas can be observed in the individuals which have HNF1A mutations even before the onset of the disease and glycosuria can be observed in the patients as well as healthy individuals with HNF1A mutant carriers (Stride et al., 2005). The expression of HNF4A gene occurs in many tissues especially the liver and the pancreas. HNF4A gene plays a role in hepatic gluconeogenesis in the liver and for glucose metabolism, the expression and secretin of insulin in the pancreas (Bartoov-Shifman et al., 2002; Rhee et al., 2003).

Numerous studies have reported that there is a linkage between the region of HNF4A located at chromosome 20q and type 2 diabetes (Weedon et al., 2004; Ji et al.,1991; Zouali et al.,1997; Love-Gregory et al.,2004).

HLA belongs to class I major Histocompatibility Complex which is found to be associated with certain immune responses, such as modulating and induction of tolerance to immunity. It has been reported to be associated with Type 1 diabetes and high blood pressure however no association was observed so far with Type 2 diabetes (Silva et al., 2016; Ilian et al., 2014). Further research has revealed that individuals having impaired glucose tolerance have higher levels of HLA-G (Solini et al., 2010). Some variants of Vascular endothelial growth factor –A (VEGF-A) may contribute to the alteration in VEGF secretion, which may result in genetic variation in VEGFA gene to the pathogenesis of type 2 diabetes (Nejla et al., 2018). It has been reported that the concentrations of VEGF-A are known to be higher in individuals with type 2 diabetes which have further impact on the plasma angiogenesis (Wieczór et al., 2015). CD2AP is found to be associated with IRS-1 gene which may provide a potential therapeutic target to treat type 2 diabetes (Zhang et al.,2018). The risk of type 2 diabetes has also been found to be associated with certain gene variants of TERF1 and ELMO1 (Mehrabzadeh et al., 2016; Zee et al., 2011). TCF7L2 is a susceptibility gene for type 2 diabetes as even common single nucleotide polymorphisms are known to be associated with the disease, variations in the gene can also cause reduction in insulin response in healthy individuals (Gloyn et al., 2009; Hattersley et al., 2007). Diet has been found to be modulated by FTO gene, lipid metabolism and hepatic glucose (Mizuno et al., 2018; Ortega-Azorín et al., 2012). Apart from these genes, certain dysfunction in the mitochondria and

mitochondrial DNA mutation has been reported to cause the progression of type 2 diabetes (Jiang et al., 2017; Julia et al., 2012).

A number of impairment in the genes of mitochondria can play a critical role in the development of age-dependent insulin resistance and leads to type 2 diabetes (Singh et al., 2011; Peterson et al., 2003). The risk of having the disease is higher with individuals having genetic alterations as compared to individuals having no alteration. At present, it has been reported that around 40 or more mutations in the mitochondria are known to be related with several mitochondrial disorders and type 2 diabetes. It has been reported that even before the onset of type 2 diabetes, the copy number of mitochondrial DNA decreases in the individuals with the disease (Lee et al., 1998). The mechanism involved in Oxidative phosphorylation in the mitochondria has also been found to be decreasing in the offspring of individuals having insulin resistant diabetic patients (Petersen et al., 2004). Aside from the mitochondrial DNA variations, numerous studies have reported that in the nuclear genes polymorphisms especially single nucleotide polymorphisms (SNP) which are in relation to the function of mitochondria are known to be associated with type 2 diabetes (Prokopenko et al., 2008). In a common disease like Type 2 diabetes which is not caused by single gene disorders, the expression of the mutant allele of the disease lies at one specific gene locus, but the disease expression depends on the contribution of various gene loci (Valsania et al., 1994). Apart from mitochondria, there are several approaches which have been laid down to study the genes involved in the development of the disease. Study of candidate genes is one approach to find and identifies the susceptibility genes of the disease (Barroso et al., 2003; Stumvoll et al., 2004). Presently, fifty or more candidate genes have been studied and repo

from across the worlds which are found to be related with type 2 diabetes. The selection and study of candidate genes are done in such a way as they are believed to be involved in certain mechanism and pathways related with glucose, insulin activity and function of the pancreatic β-cell (Bonnefond et al., 2010).

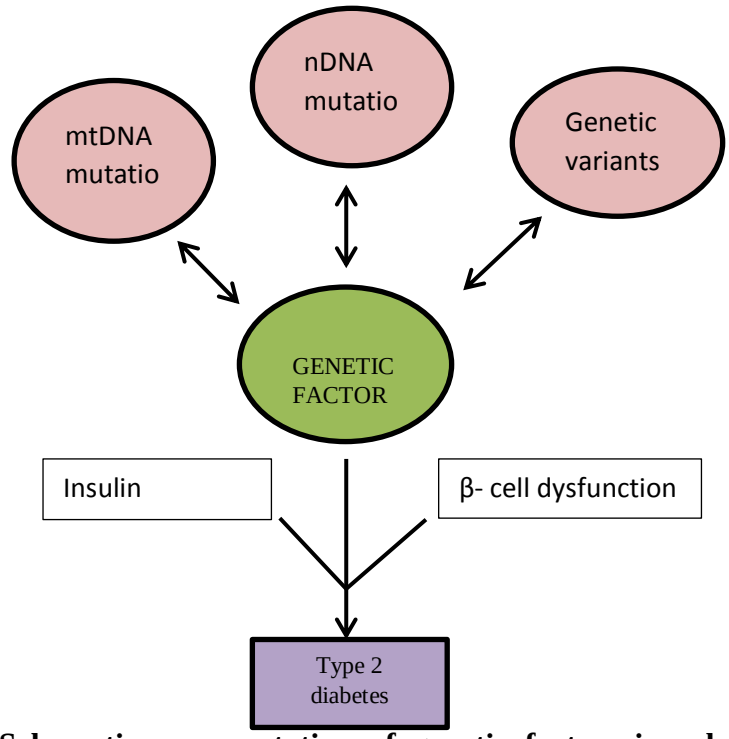


Fig 2.3: Schematic representation of genetic factors in relation with type 2 diabetes

In spite of all these studies, the results for candidate genes are sometimes contradicting. The contrasting results of different studies may be due to various reasons for like insignificant sample sizes, susceptibility in the distinction of diabetes across ethnic groups, environmental exposures, and interaction between the gene and environment (Kahn et al., 2012, Hughes et al., 2013). The development and progression of type 2 diabetes cannot be attributed to genetic factors alone and unveiling of the metabolic disorder can be understood with the help of other factors.

Mizoram belongs to one of the eight sister states of the North-eastern region of India, bordered by Bangladesh and Myanmar and is the second least populous state in the country (RGCCS, 2011). Mizoram has shown an increase in reports and cases of type 2 diabetes mellitus, however no in depth records have been found to fully understand the occurrence of the disease. Therefore, it becomes a huge challenge to find the factors which may contribute in the occurrence of type 2 diabetes in a population where the prevalence of the disease is high.

Chapter III

Aim and Objectives

Aims and Objectives

The following aims are set forth to carry out the proposed work in this study:

- Study of mutations related with candidate genes and their association with type 2 diabetes in Mizo Population.
- To study the demographic factors and clinico-biochemical profiles of type 2 diabetes patients.

Chapter IV

Materials and Methods

4. Materials and Methods

Participants

Five hundred previously diagnosed T2DM patients receiving care from major diagnostic and health clinic 'Genesis Laboratory', Aizawl, Mizoram and five hundred controls volunteered to participate in this study. The age ranges for both the cases and controls were between 40 - 85 years were included in this study. Both the cases and controls included for this study belonged only to Mizo ethnic group. The work has been approved by ethical committees of Civil Hospital, Aizawl (B.12018/1/13-CH(A)IEC/39 dtd. 23/12/2015) and Human Ethical Committee, Mizoram University (MZU/IHEC/2015/006 dtd. 14/12/15).

Inclusion criteria

- The selection of cases was done after the disease was confirmed by a diabetologist.
- Fasting plasma glucose levels >120 milligram per deciliter (mg/dl).
- Post prandial plasma glucose levels >200 milligram per deciliter (mg/dl).
- Control group comprised of non-diabetic subjects without the symptoms of diabetes mellitus, blood glucose levels within normal limits (below < 120 mg/dl).

Exclusion criteria

- Gestational diabetes
- Diabetes due to pancreatic disorder
- Pregnancy and Other types of severe illness and metabolic diseases

4.1. Study on the demographic factors and clinico-biochemical profiles of type 2 diabetes patients.

4.1.1. Study Design and assessment

The epidemiological information was collected with consent of the participants. Information on demographic factors, dietary habits, tobacco, alcohol use and family history were recorded. Dietary habits like meat, smoked meat, salt intake and sa-um were taken into account. Lifestyle habits such as cigarette, sahdah, tuibur, paan and alcohol consumption were recorded. Dietary habits were grouped into three categories viz., little (one to two days per week), moderate (two to three days per week), high (more than three days per week). Blood sugar and cholesterol levels were estimated by GOD – POD method using Sys 200 biochemistry analyzer (Shrestha et al.,2008; Trinder et al.,1969). Estimation of creatinine was done by the modified Jaffe's method by using Sys 200 biochemistry analyzer (Shrestha et al.,2008; Bowers et al.,1980). The types of medication taken by the patients were recorded as oral or injection. History of hypertension was also taken into account. BMI was calculated as weight (kg)/height^2 (m2) and used to categorize BMI-measured weight status: underweight (BMI $\leq$ 18.5), normal (BMI 18.5–24.9), overweight (BMI 25.0–29.9) and obese (BMI $\geq$30) (WHO, 1999; WHO, 2000).

4.1.2. Statistical Analysis

For the different factors from diabetic patients and healthy controls with age range 40-85 years (mean age range =65.5) were analyzed. All statistical analysis was performed using SPSS 20.0 version (IBM corp, Armonk, NY) software package. Chi-square tests were used to assess the association between demographic factors and type 2 diabetes. Factors that were deemed of potential importance to the

univariate analysis ($P<0.05$) were included in the multivariate analysis using Cox regression (Karen et al., 2003). Logistic regression analyses were carried out to calculate the influence of lifestyle and dietary factors for Type 2 Diabetes. The independent effect of risk factors was investigated in a multivariate model (introducing all variables and terms of interactions) retaining only statistically significant or factors showing a confounding effect only. Gender and various life style and food habits were all considered in the regression model as potential confounders to evaluate the association of risk factors and susceptibility to type 2 diabetes. The variables were adjusted with Age. ROC curve was plotted for the factors which achieved high odd ratio to estimate the potential risk score. Unpaired t-test was performed for male and female diabetic patients to find the difference in the levels of biochemical tests.

4.2. Evaluation of total protein, amylase and microalbumin

This study included three (3) groups of diabetic patients, 50 cases newly diagnosed, 50 cases 1-2 years diagnosed and 50 cases more than 3 years type II diabetes mellitus (T2DM). These groups were matched with age matched healthy individuals of 50 controls. A total of 200 individuals with age between 30 to 70 participated in the study. The participants were taken into account that they belong only to Mizo ethnic tribe.

Inclusion criteria

As described above

Exclusion criteria

Patients with severe diabetic complications, other serious illness, habits of smoking and alcohol, complications of the head and neck region, and individuals with any oral lesions were excluded from the study.

The study protocol was thoroughly explained and informed consent was obtained from the patients. Samples from the blood and saliva were collected from each participant.

4.2.1. Samples:

(a) Saliva and serum

5.0 ml of fasting whole blood was collected from the patients and healthy individual, stored for about 1 hr. in a plain vacutainer at room temperature. The tube was centrifuged (3000×g) for 10 minutes, the clear serum was pipetted into clear dry tube and then stored at (-20) °C for subsequent analysis. Fasting whole saliva is collected using passive drool method in a polypropylene tube, after thoroughly rinsing the mouth with saline solution (Jacobs et al., 2005). It was centrifuged (3000×g) for 15 minutes and the supernatant was stored at (-20) °C for further analysis.

(b) Urine

Fasting urine samples was collected (~3 to 5 ml) in a clean and sterile urine collection container for further investigation.

4.2.2. Total protein determination:

Total protein estimation was performed using Bradford method where Bovine serum albumin (BSA) was used as the standard protein (Fanglian et al., 2011). Various concentrations of standard protein (0.2, 0.4, 0.6, 0.8 and 1.0 ml) were taken and the volume was made up to 1 ml. Further, 0.2ml of sample (saliva/serum) was taken in a test tube and the volume was made up to 1 ml. To this, 5 ml of coomassie brilliant blue was added and the solution was mixed by vortexing. It was then incubated at room temperature for about 10 to 30 minutes. The absorbance was read at 595 nm using a spectrophotometer against a suitable blank. The Values obtained were expressed as mg/dl.

4.2.3. Estimation of Amylase Activity:

Amylase activity was performed using Amylase enzyme activity assay where starch was used as the substrate (Biochemden, 2018). 0.5 ml of substrate (starch) and 0.2 ml of 1%Nacl were taken in a clean and dry test tube and pre-incubated at 37°C for 10 minutes. 0.3 ml (saliva/serum) was added and incubated at 37°C for 15 minutes and 1 ml of DNS (Dinitro Salicylic acid) was added and kept in boiling water for about 10 minutes. It was then cooled and diluted with 10 ml of distilled water. The colour developed was read at 520 nm using spectrophotometer using a suitable blank. The Values obtained were expressed as mg/dl.

The Mean absorbance change per minute for protein estimation and amylase activity was calculated and expressed as units per liter (López et al.,2003; Rantonen et al.,2000; Henskens et al.,1996) .

Urine Microalbumin:

10-20 µl urine samples were taken and TruCal Albumin U/CSF (Diasys) kit was used and was run in Sys 200 chemistry Analyser for microalbumin estimation.

4.2.4. Statistical Analysis

All statistical analysis were performed using SPSS 20.0 version,(IBM corp, Armonk, NY) software package. Values were expressed as means ± standard deviation and a $P \leq 0.05$ was considered significant. Intra group correlations of amylase, and total protein were done by using ANOVA. For Urine microalbumin, Paired t-Test was performed to find the intra group correlation.

4.3. Study of mutations related with candidate genes and their association with type 2 diabetes in Mizo Population.

Blood DNA isolation

Lymphocytes from whole blood was separated by lysing the red blood cells [RBCs] using a hypotonic buffer with minimal lysing effect on lymphocytes. Lymphocytes was digested with 100 mg/mL proteinase K in cell lysis buffer for 2-3 hours at 56°C and extracted twice with phenol : chloroform : Isoamyl alcohol [25:24:1]. The DNA was precipitated with two volumes of chilled isopropanol, and 1/10 volume of 3 M sodium acetate. The precipitated DNA was washed by 70% ethanol and the DNA was eluted by nuclease free water or TE buffer (Ghatak et al., 2013).

The mutational studies were carried out using the following approaches:

4.3.1. Study of candidate genes using Sanger sequencing

The candidate genes (PPARG, KCNJ11, HNF4A and HNF1A) were amplified using their specific primers (Table 4.1). Polymerase chain reaction [PCR] was carried out in 25 µL total reaction volumes, each containing 100 ng of template DNA, 0.2 pM of each primer, 2.5 µL of 10X PCR buffer, 1.5 mM $MgCl_2$, 0.2 mM dNTPs, and 1 U of DNA polymerase. PCR was performed using the following conditions, initial denaturation at 95°C (5 mins), denaturation at 95°C for 40 secs, annealing for 40 secs at 62°C for KCNJ11 and HNF1A gene, 55°C for PPARG gene and 60°C for HNF4A genes and extension at 72°C (40secs) with final extension at 72°C (5 mins). The amplified PCR product was checked in 1.2% agarose gel and visualized using Bio-rad Gel documentation system.

GENE	EXON	PRIMER (5'-3')	PRODUCT SIZE (bp)
KCNJ11	2	GAGGAGTGCCCACTGGCC CACAGAGTAACGTCCGTCCTCC	567
PPARG	3	GACCATGGTTGACACAGAGA GTACTCTTGAAGTTTCAGGTC	220
HNF4A	2	GCCAAGGGGCTGAGCGAT CTCCCAGCAGCATCTCCTG	235
HNF1A	1	GCAGACGGAGCTCCTGGC CACGGCTTTCTGGTGGGC	286

Table 4.1: Candidate gene primers and their associated exons

Sequencing and Sequence Analysis

Based on the PCR results, the samples were sequenced from the opposite direction to ensure reading accuracy. Sequences and chromatograms obtained was examined by chromas software version 2.13, DNA baser and align by BLAST

[www.ncbi.nlm.nih.gov/blast]. The results of the DNA sequence analysis was compared with the published Cambridge Sequence using Mutation Surveyor Version 1.4 DNA mutation analysis software [Softgenetics, State College, PA]. The exons of the candidate genes was checked from Ensembl, HGNC and Entrez data base and protein features and phenotypic effect was checked from Uniprot and OMIM data base respectively, as shown in table 4.2.

GENE	HGNC	ENTREZ	ENSEMBL	OMIM	UNIPROT KB
KCNJ11	6257	3767	ENSG00000187486	600937	Q14654
PPARG	9236	5468	ENSG00000132170	601487	P37231
HNF4A	5024	3172	ENSG00000101076	600281	P41235
HNF1A	11621	6927	ENSG00000135100	142410	P20823

Table 4.2: Ensembl, HGNC, Entrez Uniprot and OMIM ID's of the candidate genes from public data base.

All the sequences containing the mutation was evaluated for their potential pathogenicity using the following softwares: DNA baser version 3.5.4.2, Codon Code aligner version V.4.2.2, Mutation taster [www.mutationtaster.org/], PolyPhen-2 [http://genetics.bwh.harvard.edu/pph], SIFT [http://sift.jcvi.org], Mutation Assessor [http://mutationassessor.org/]. The result of PolyPhen-2 was retrieved from the original webpage [version 2.2.2] which were used for weighted average scores.

4.3.2. Study of candidate genes- Case Control Association study using SNP genotyping

The candidate genes were studied using SNP genotyping methods involving Tagman Allele discrimination Assay and Agena Mass Array.

4.3.1. Tagman Allele discrimination Assay

DNA isolated from 425 cases and 330 healthy controls were used for the study. Quantification of the DNA was performed using Bio-spectrophotometer (Eppendorf, Hamburg, Germany) and the working dilutions were made (5 ng/µl) for genotyping.

Genotyping

The genotyping of ***UCP3*** variant rs1800849 and ***MACF1*** rs2296172 were performed using allele discrimination assay on Real time PCR make (Mx3005P Agilent USA). UNG Master Mix (Applied Biosystem, USA) and Taqman assay (Predesigned Primer and Probe, labelled with Fam and VIC supplied by Applied Biosystem, USA) was used for genotyping. As recommended by the manufacturer, dilution of the assay mix was made from 40X concentration to 20X with TE (Tris-EDTA) buffer. PCR reactions were carried out in 96-well plate format with three negative controls (NTC) to check for extraneous nucleic acid contamination. The volume of the total PCR reaction mix in each well was 10 µl, contributed by 2.5 µl of Taqman UNG master mix, 0.25µl of 20X assay, 3µl DNA (5 ng/µl) and 4.25 µl of water added to make up the final volume. The PCR conditions were: hold for 10 minutes at 95°C, then 40 cycles of 95°C for 15 seconds and 60°C for 1 minute. The post PCR detection system (Mx3005P Agilent USA) is used to measure allele specific fluorescence and alleles were called automatically. Ninety-three random diabetic samples were picked and re-genotyped for cross validation; genotyping calls were observed with 98 percent concordance.

Statistical analyses

After genotyping, statistical analysis of the data was done using SPSS software (version 23; Chicago, IL). Chi square ($\chi 2$) analysis was performed and genotypic frequencies were tested for the Hardy- Weinberg equilibrium. The odds ratio (OR), confidence interval (CI) and level of significance (p-value) were also estimated by performing logistic regression analysis and corrected for potential confounding factors like age, gender and Body mass index (BMI). Population Attributed Risk (PAR) percentage was also calculated, with 95% CIs, by using adjusted OR. The power of the study was calculated by PS software version 3.1 (WD DWaP., 1997).

4.3.2. Agena Mass Array

Isolated DNA from 384 T2D cases and 384 healthy controls were used for the study. Quantification of the DNA was performed using Biospectrophotometer (Eppendorf, Hamburg, Germany) and the working dilutions were made (10 ng/µl) for genotyping.

Variant selection and protocol

SNPs selections in the present study (Table 4.3) were included from the genes which have been implicated with T2D or the traits associated with T2D. All of the variants which selected for the study have never been evaluated in Mizo population. Genotyping of samples was performed using Agena Mass array platform in Human Genetic Research Group at Shri Mata Vaishno Devi University, Katra, Jammu, India. Customized forward, reverse and single base extension primers were designed using online Agena cx platform V.2.0 (www.agenacx.com). To detect a variation initially targeted region was amplified by using multiplex PCR. One µl of genomic DNA

(concentration of 10 ng/ul) was loaded in 384 well PCR plates which were dried at 85^{o}c for 10 minutes. After drying, the reaction mixture was prepared containing dNTPs, primers pool (forward & reverse), reaction buffer and DNA polymerase. After completion of first PCR, the reaction was treated with shrimp alkaline phosphatase (SAP), the multiple PCR reaction then subjected to single base extension reaction using mass modified ddNTPs and primers (pooled single extension primers). PCR conditions were adopted from Gabriel et al. (2009). Further, the final PCR product was treated with cationic resin and then energy transferred to spectro-chip. The transferred product then fired to MT analyzer. The data was then processed and analyzed by preinstalled Typer Analyzer v.4.0.

S.No.	**SNP**	**Chromosome: position**	**Location of the Variant w.r.t Gene**	**Candidate Gene and its Functional role**
1.	rs16847897	3:169850328	upstream to *TERC*; intron variant of *LRRC31*	*TERC*: Encodes for non-coding RNA that provides template for telomere replication
2.	rs1063320	6: 29830972	3 Prime UTR Variant	*HLA-G:*Gene Ontology (GO) annotations related to this gene include *protein homodimerization activity* and *peptide antigen binding*.
3.	rs2010963	6:43770613	5 Prime UTR Variant	*VEGFA*: It encodes a heparin-binding protein, which exists as a disulfide-linked homodimer. This growth factor induces proliferation and migration of vascular endothelial cells, and is essential for both physiological and pathological angiogenesis.
3.	rs3025020	6:43781373	Intron Variant	VEGFA:It encodes a heparin-binding protein, which exists as a disulfide-linked homodimer. This growth factor induces proliferation and migration of vascular endothelial cells, and is essential for both physiological and pathological angiogenesis.
4.	rs9369717	6:47586732	Intron Variant	CD2AP: encodes a scaffolding molecule that regulates the actin cytoskeleton.
	rs9349417	6:47612921	Intron Variant	
5.	rs741301	7:36878390	Intron Variant	*ELMO1*: Encodes a member of the engulfment and

	rs1882080	7:36916236	Intron Variant	cell motility protein family. These proteins interact with dedicator of cytokinesis proteins to promote phagocytosis and cell migration.
6.	rs6982126	8:73027388	Intron Variant	*TERF1*: Encodes a telomere specific protein which is a component of the telomere nucleoprotein complex
7.	rs290475	10:113114260	Intron Variant	TCF7L2: Encodes a high mobility group (HMG) box-containing transcription factor that plays a key role in the Wnt signaling pathway. The protein has been implicated in blood glucose homeostasis. Genetic variants of this gene are associated with increased risk of type 2 diabetes.
8.	rs9939609	16:53786615	Intron Variant	FTO: This gene is a nuclear protein of the AlkB related non-heme iron and 2-oxoglutarate-dependent oxygenase superfamily but the exact physiological function of this gene is not known.
9.	rs2853826	MT:10398	Missense Variant	*ND3:* Gene Ontology (GO) annotations related to this gene include *NADH dehydrogenase (ubiquinone) activity*.

Table 4.3 : Candidate Genes selected for the study and their functional role.

Statistical analysis

The statistical analysis were mainly performed using Plink V.1.0962 with maximum 10,000 permutations. Each SNP was tested for Hardy Weinberg Equilibrium (H.W.E). A significant association of SNPs was evaluated by 3×2 chi square tests for genotypic frequencies between T2D cases and controls. Further logistic regression analysis was performed for dominant model using SPSS V.23 in order to obtain corrected odds ratio (OR), confidence interval (CI) and p-value as level of significance from confounding factors like age, gender and BMI.

4.3.3. Study of candidate genes using whole exome sequencing

Four (04) diabetic samples with strong familial history (02 Males and 02 Females) and 6 controls were taken and analysed.

Sequencing

The first step includes fragmentation of DNA where 100 ng of genomic DNA was fragmented to a 150 bp insert size with 3' or 5' overhangs using Covaris. The shearing buffer contains 5 ml of Resuspension buffer (RSB) and 10 ml of EDTA in a 15ml conical flask. 50 µl of DNA was transferred to separate the covaris well plates. Sample purification beads (SPB) were vortexed thoroughly until it was well dispersed and washing was done using 80% ethanol. This process converts the overhangs resulting from fragmentation into blunt ends using ERP3 (End Repair Mix). The 3' to 5' exonuclease activity of this mix removes the 3' overhangs and the 5' to 3' polymerase activity fills in the 5' overhangs. Following end repair, the library size is selected using SPB (Sample Purification Beads). A single 'A' nucleotide is added to the 3' ends of the blunt fragments to prevent them from ligating to each other during the adapter ligation reaction. A corresponding single 'T' nucleotide on the 3' end of the adapter provides a complementary overhang for ligating the adapter to the fragment. This strategy ensures a low rate of chimera (concatenated template) formation. The next process ligates multiple indexing adapters to the ends of the DNA fragments, which prepares them for hybridization onto a flow cell. PCR was used to selectively enrich those DNA fragments that have adapter molecules on both ends and to amplify the amount of DNA in the library. PCR is performed with PPC (PCR Primer Cocktail) that anneals to the ends of the adapters. Probes were hybridized which combines DNA libraries containing unique indexes into a single pool, and then binds targeted regions of the DNA with capture probes. The hybridized probes were then captured using SMB (Streptavidin Magnetic Beads) the targeted regions of interest. Two heated washes remove nonspecific binding from the

beads. The enriched library was eluted from the beads and prepared for a second round of hybridization. Second hybridization step was performed by binding the targeted regions of the enriched DNA with capture probes. This second hybridization ensures high specificity of the captured regions. Second capture was performed using SMB (Streptavidin Magnetic Beads). Cleaning up of the captured library was done using SPB (Sample Purification Beads) to purify the captured library before PCR amplification. 8-cycle PCR program was performed to amplify the enriched library. The enriched library was then validated by running 1 µl of post enriched library on an Agilent Technologies 2100 Bioanalyzer using a High Sensitivity DNA chip. The library preparation was done using the TruSeq Exome Library Prep Reference Guide (Illumina).

Sequence Analysis

FASTQ files from sequencing data generated was analyzed for quality control where the quality of the reads were checked, filtered, trimmed and the adapters were removed using FastQC/ Trimmomatic software (Andrews, 2010; Bolger, 2014). The filtered FastQ files were then mapped against the reference genome, in this case GRCh 37-hg19 genome assembly using Stampy/BWA Stampy hybrid **(**Lunter and Goodson, 2011; Li and Durbin, 2009) which in turn gives the SAM files. The sorting out, marking of the duplicates and indexing of SAM files into BAM files was done using Samtools/Picard (Li, 2009; http://broadinstitute.github.io/). From the BAM files, variant calling which may consists of detection of SNP's and INDELS was done using Platypus (https://www.well.ox.ac.uk/platypus)which gives rise to the Variant calling file

(VCF). From the variant calling file, the variants were annotated using ANNOVAR (Wang, 2010). The annotated variants were filtered using only pathogenic/disease causing scores form the prediction software's like SIFT (https://sift.bii.a-star.edu.sg/), Polyphen2 (gene tics.bwh.harvard.edu/), LRT SCORE (http://www.genetics.wustl.edu/), LR (Alan et al., 2007), FATHMM (fathmm.biocompute.org.uk/), MUTATIONTASTER (www.mutationtaster.org/) and MUTATIONACCESSOR (mutationassessor.org/). The analysis from the annotated variants was also carried out using panel genes consisting of 582 diabetic genes from GWAS catalogue. The genes included other diabetic related diseases like diabetic neuropathy, nephropathy etc. The prediction of the effect of substitutions between amino acids based on chemical properties were done using Grantham matrix score (www.mutationtaster.org/) and the effect of certain mutation on the protein structure was predicted using HOPE (http://www.cmbi.ru.nl/hope/) and the possible pathways in which the unreported variants may belong was analyzed using KEGG pathway. To further validate the prediction by other database, i mutant 2.0 have been used to predict the changes in protein stability upon mutation from the protein sequence (http://folding.biofold.org/i-mutant/i-mutant2.0.html).

4.3.4. Study of Candidate genes using whole mitochondrial sequencing

28 diabetic DNA samples with familial diabetic history and 12 controls were taken for the study.

Sequencing of mtDNA

Equimolar pools of gel purified amplicons were used for the preparation of sequencing libraries using Nextera XT DNA library kit (Illumina). Libraries were pooled and sequenced in Illumina HiSeq-2500 to generate 2 x 100 bp reads.

Fragmented products were purified with Ampure XP (Beckman Coulter). Library preparation was performed using Ion Plus Fragment Library kit (Thermo Fisher Scientific) following the manufacturer's protocol. The quality and quantity of the libraries were assessed by High Sensitivity DNA chip (Agilent) in 2100 Bioanalyzer (Agilent) and Picogreen dye in Qubit Fluorometer (Invitrogen). All the mtDNA sequencing libraries were pooled and sequenced in Ion PGM sequencer (Thermo Fisher Scientific) using Ion 318TM chips (Thermo Fisher Scientific).

Sequence Analysis

FASTQ files from mitochondrial amplicon sequencing data generated in Illumina HiSeq-2500 were analyzed using FASTQC for preliminary quality checking. BWAMEM (Li et al., 2009) was used to align the sequence data to the mitochondrial reference sequence (rCRS- revised Cambridge Reference Sequence). The SAM file thus produced was converted to BAM files using SAM tools (García-Alcalde et al., 2012) and reads below mapping quality of 40 were discarded. BAM files were quality checked using QualiMap (Koboldt et al., 2012) to ensure sufficient sequence coverage along the entire mtDNA reference sequence. VarScan2 (Homer) was used to detect germline variants in the mtDNA. Variants with a base quality score (q) less than 20 were removed. Variant calls that were not supported by at least 3% of reads or exhibited less than 10% of total reads in either direction were removed. Briefly, sequence reads were aligned to rCRS by TMAP (R Core Team,2014). mtDNA mutations in blood samples were identified by the plug–in variant caller. Visualization and annotation of variants were done using Integrative Genome Viewer (Wang et al., 2010) and ANNOVAR (Kloss-Brandstätter et al., 2011), respectively. From the annotated data, the specificity of the mutation was

analyzed using the mitochondrial genome databases like MITOMAP (http://www.mitomap.org), mtDB (http://www.genpat.uu.se/mtDB/), mtSNP (http://www.mtsnp.tmig.or.jp/_mtsnp/index_e.shtml). MITOMASTER was used to identify nucleotide variants relative to the rCRS, and to view species conservation (https://www.mitomap.org/foswiki/bin/view/MITOMASTER/WebHome). CIRCOS (Krzywinski, M. et al.2009) and Heatmap (https://biit.cs.ut.ee/clustvis/) were plotted to represent the relationship between the variants for cases and controls. The hydrophobic / hydrophilic natures of mitochondrial proteins were predicted by the SOSUI system (http://sosui.proteome.bio.tuat.ac.jp) (Guo et al., 2005). Fisher exact test was used to calculate the frequency of the alleles between the cases and controls. Odds ratios (OR) with 95% of confidence intervals (CI) were calculated to estimate strength of association. p values less than 0.05 were considered statistically significant. All statistical analyses were performed using the SPSS 20.0 version, (IBM corp, Armonk, NY). Potential impacts of non-synonymous substitutions on proteins were predicted using Pmut (http://www.ics.uci.edu/~baldig/mutation.html) (Cheng et al., 2006).

Chapter V

Results

5. Results

5.1. Study on the demographic factors and clinico-biochemical profiles of type 2 diabetes patients

From the total of 1000 participants, 500 were diabetic [Males - 261 (52.3%) and Females - 239 (47.7%)] and 500 were healthy control [Males - 238 (47.7%) and Females - 262 (52.3%)]. The potential risk factors in univariate analysis using chi-square test were alcohol, sahdah, paan, tuibur, meat, saum, smoked meat and salt (Table 5.1) and these were further analyzed using multivariate analysis. The risk of type 2 diabetes was observed to be higher in patients who consumed Saum which is the fermented pork fat (Odd ratio, OR: 18.98, 95% Confidence interval, 95% CI: 9.8182 - 36.6918). This was observed with the adjusted age of 45 years or more for both males and females (OR: 0.0893, 95% CI: 0.0419 - 0.5300). Other life style risk factors that were further analyzed includes betel leaves with areca nut chewing (OR: 0.1006, 95% CI: 0.0537 - 0.1885), tuibur (OR: 0.1243, 95% CI: 0.0530 - 0.2918) and dietary habits like smoked meat (OR: 0.0703, 95% CI: 0.0412 - 0.1200). Consumption of salt in excess may also be a risk for Type 2 diabetes (OR: 0.2134, 95% CI: 0.1400 - 0.3251) as described in Table 5.2.

Factors	Univariate Analysis		
	Odd Ratio	**95% CI**	**P value**
Gender	1.024	0.799 - 1.313	0.8493
Age in years (>45 Control vs >45 Diabetic)	0.0893	0.0426 - 0.1871	<0.0001
Cigarette Smoking	0.7465	0.5940 - 0.9382	0.0122
Alcohol Consumption	0.0117	0.0016 - 0.0844	<0.0001
Paan Consumption	0.0571	0.0348 - 0.0939	<0.0001
Sahdah Consumption	1.8776	1.5139 - 2.3286	<0.0001
Tuibur Consumption	2.0130	1.3908 - 2.9137	0.0002
Saum Consumption	2.2667	1.7115 - 3.0019	<0.0001
Meat Consumption	0.4115	0.3344 – 0.5064	<0.0001
Smoked Meat Consumption	0.0695	0.0499 - 0.0970	<0.0001
Salt Consumption	0.2882	0.2309 - 0.3597	<0.0001

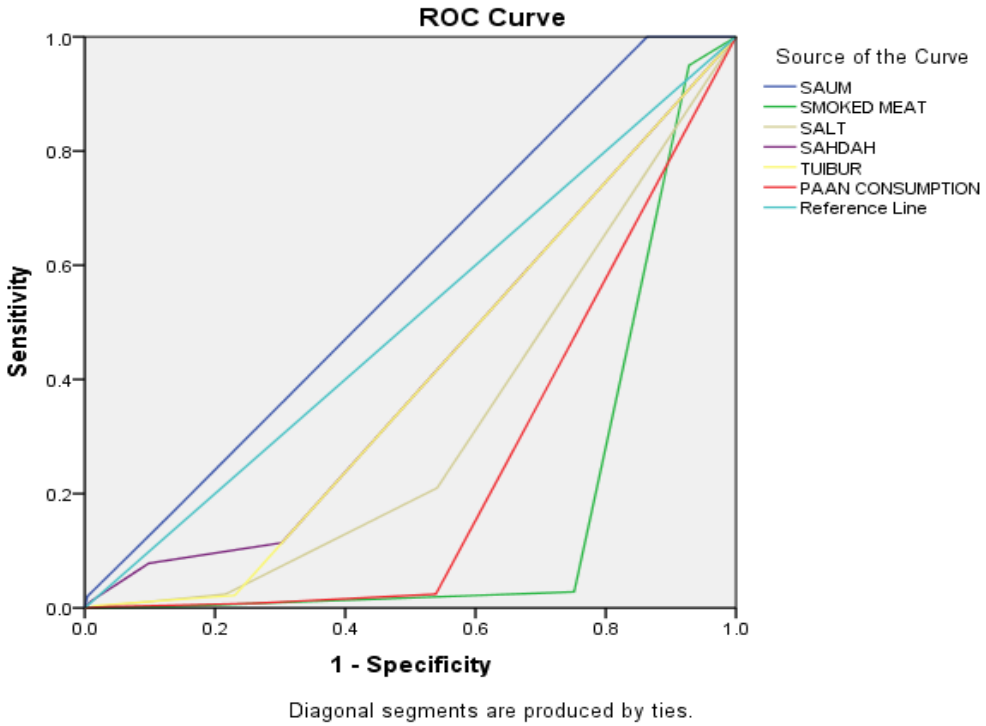
Table 5.1: Univariate analysis of dietary and lifestyle factors

	Multivariate Analysis		
Factors	Odd Ratio	95% CI	P value
Age in years (>45 Control vs >45 Diabetic)	0.1491	0.0419 - 0.5300	0.0033
Cigarette Smoking	1.3529	0.8268 - 2.2137	0.2289
Alcohol Consumption	0.0573	0.0068 - 0.4793	0.0083
Paan Consumption	0.1006	0.0537 - 0.1885	<0.0001
Sahdah Consumption	1.0003	0.6869 - 1.4566	0.9987
Tuibur Consumption	0.1243	0.0530 - 0.2918	<0.0001
Saum Consumption	18.9802	9.8182 - 36.6918	<0.0001
Meat Consumption	0.9583	0.6481 - 1.4172	0.8312
Smoked Meat Consumption	0.0703	0.0412 - 0.1200	<0.0001
Salt Consumption	0.2134	0.1400 - 0.3251	<0.0001

Table 5.2.: Multivariate analysis of dietary and lifestyle factors

ROC curve was plotted to find the specificity (89.18) and Sensitivity (96.20) of the results from the multivariate analysis, where the area under the ROC curve was observed to be 0.947 (95% CI = 0.932 to 0.960) and P <0.0001 which has been predicted to be a potential hazard score for these factors (Figure 5.1). Saum consumption acts as a potential risk factor after performing the cox-PH model. Smoked meat, excess salt, tuibur and paan consumption acts as confounding factors for type 2 diabetes. No correlation was found between BMI and type 2 diabetes in Mizo population.

Between the male and female diabetic patients, we observed no difference in biochemical parameters like fasting glucose level (p=0.9813), post prandial glucose level (p=0.9148), cholesterol (p=0.5673) and HbA1c (p=0.0839). However, creatinine level (p=0.0382) was observed to be differentially regulated between male and female diabetes patients in Mizo population (Figure 5.2).



Area Under the Curve	
Test Result Variable(s)	**Area**
SAUM	**0.947**
SMOKED MEAT	0.166
SALT	0.318
SAHDAH	0.412
TUIBUR	0.396
PAAN CONSUMPTION	0.242

Figure 5.1: Estimated ROC curve for the significant demographic factors in multivariate analysis.

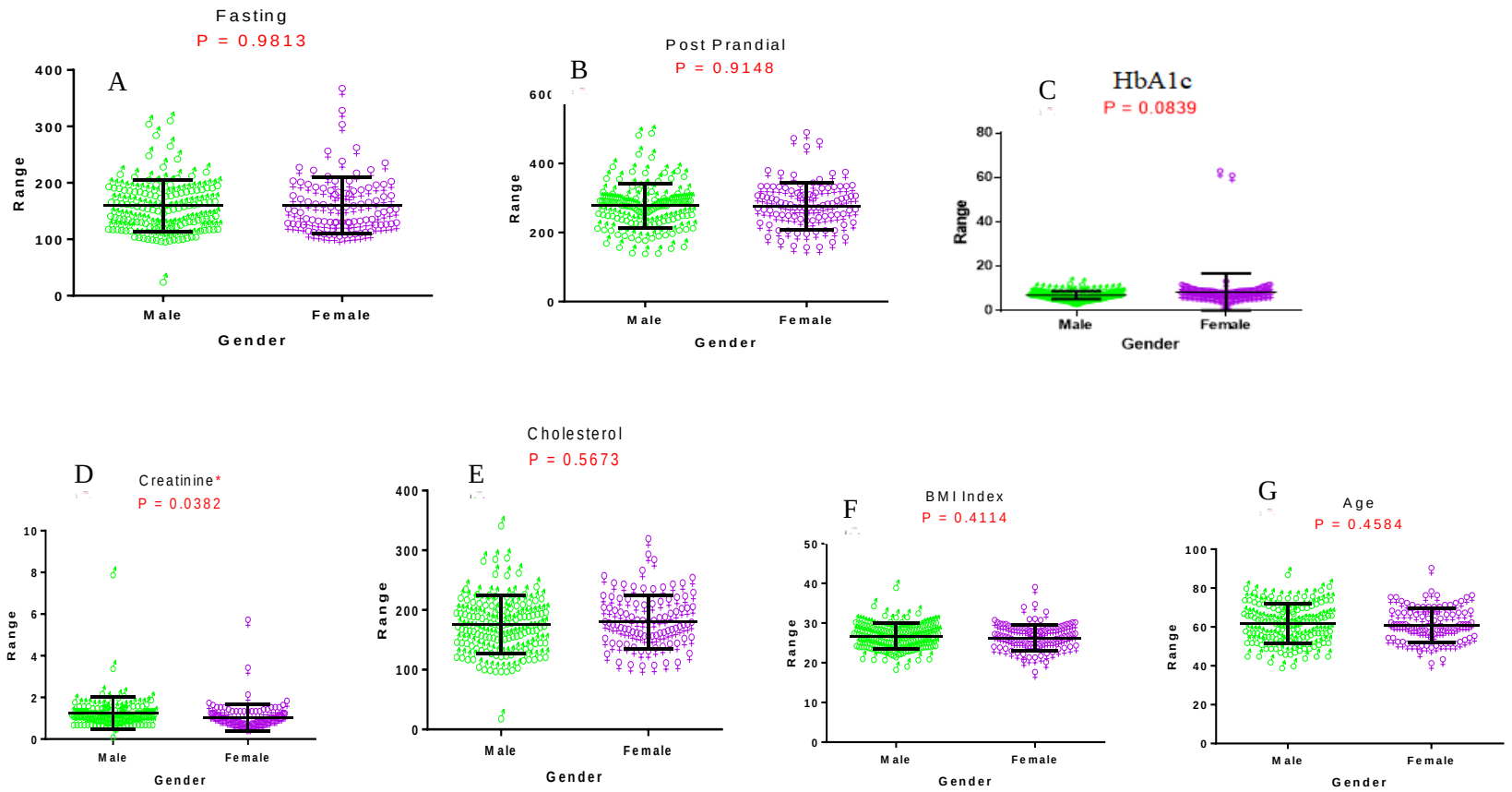


Figure 5.2: Scatter Plot showing comparison of different clinical factors (after performing unpaired t-test) between male and female diabetic patients. A: Fasting blood sugar, B: Post Prandial blood sugar, C: HbA1c, D: Creatinine, E: Cholesterol, F: BMI index, G: Age.

55.2. Evaluation of total protein, amylase and microalbumin

Evaluation of total protein, amylase and microalbumin were performed in 200 patients [50 cases newly diagnosed (Group I), 50 cases 1-2 years diagnosed (Group II), 50 cases more than 3 years (Group III) type 2 diabetes mellitus (T2DM) and 50 healthy controls (Group IV)]. The mean age of the study group was observed to be 50 ± 5 years and the salivary pH was between 6 to 7. The result obtained from the total protein concentration and amylase activity from both the serum and saliva were compared between the groups. All the analysis was done using ANOVA.

We observed that the mean salivary protein level was found to be lower in Group I (0.043±0.038), slight increase in Group II (0.058±0.048), and the level of protein dropped in Group III (0.022±0.022), when compared to control group IV (0.058±0.016) which showed a statistical significance (P < 0.0001 and F-value 10.29) (Table 5.3; Figure 5.3).

GROUPS	N	MEAN	SD	P	F-VALUE
CONTROL	50	0.058	0.016		10.290
I. NEWLY DIAGNOSED	50	0.043	0.038	0.01	
II. 1-2 YEARS DIAGNOSED	50	0.058	0.048	0.01	
III. MORE THAN 3 YEARS	50	0.022	0.022	0.01	

Table 5.3: Comparison of type 2 diabetes mellitus and control groups in relation to salivary protein (mg/dl) by ANOVA

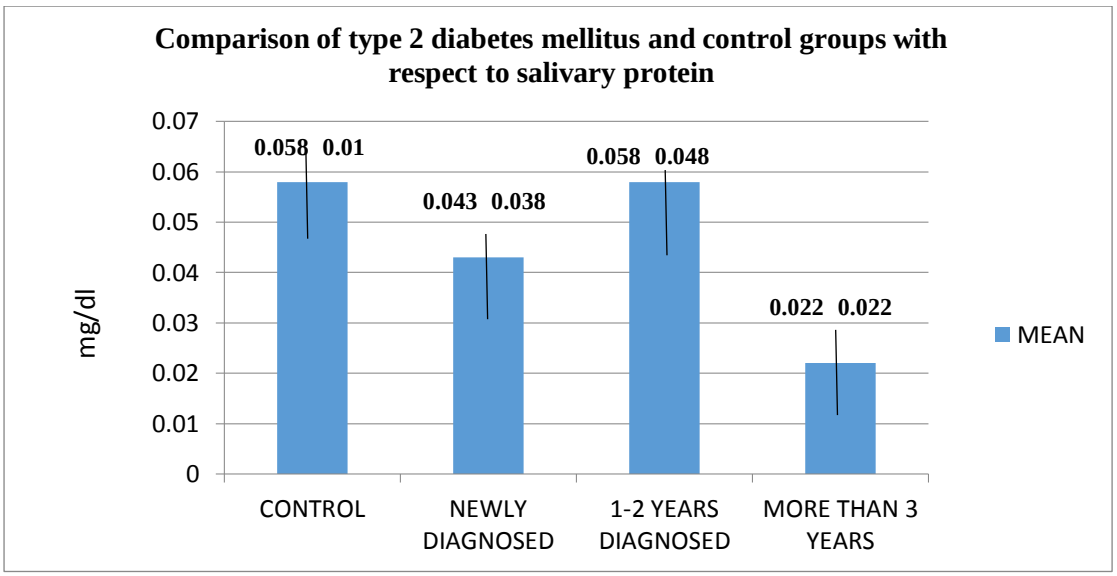


Figure 5.3: Graphical representation of Type 2 diabetes mellitus and control groups in relation to salivary protein levels (mean± standard error)

The mean serum protein level was also found to be lower in the patient diabetic groups, newly diagnosed (0.089±0.016), the protein level slightly rise in 1-2 years diagnosed (0.106±0.010), and it dropped in more than 3 years diagnosed (0.076±0.016) when compared to control groups (0.117±0.023) which showed a statistical significance with F-value of 55.857 and P < 0.0001 (Table 5.4; Figure 5.4).

GROUPS	N	MEAN	SD	P	F-VALUE
CONTROL	50	0.117	0.023		55.857
I.NEWLY DIAGNOSED	50	0.089	0.016	0.00	
II.1-2 YEARS DIAGNOSED	50	0.106	0.010	0.006	
III.MORE THAN 3 YEARS	50	0.076	0.016	0.000	

Table 5.4: Comparison of type 2 diabetes mellitus and control groups in relation to serum protein (mg/dl) by ANOVA

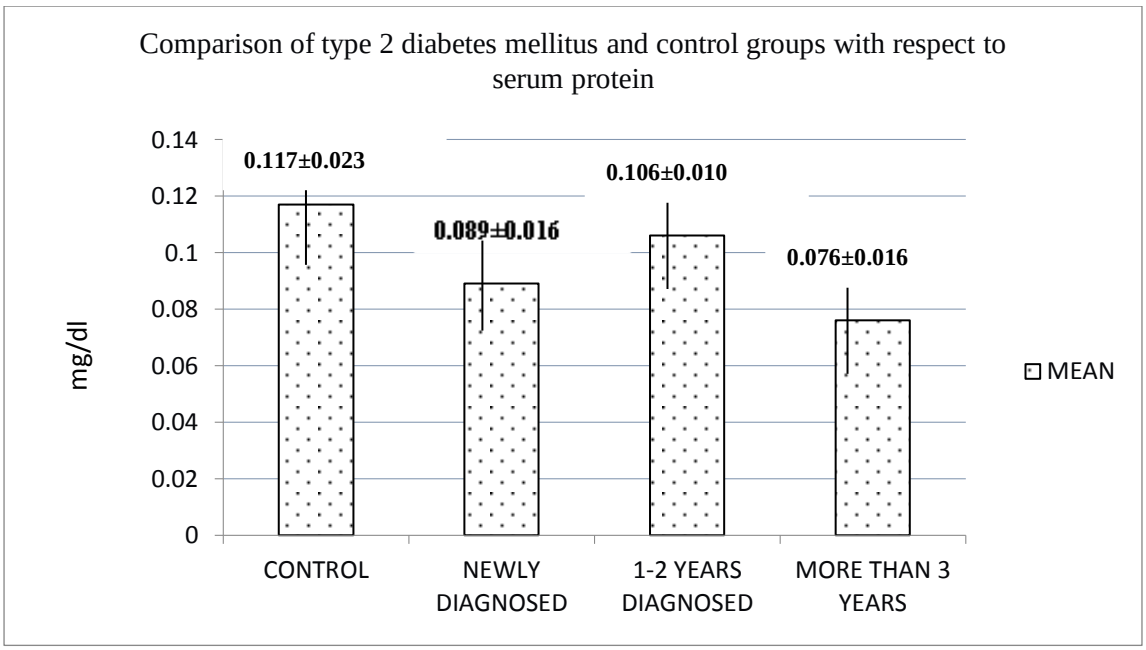


Figure 5.4: Graphical representation of Type 2 diabetes mellitus and control groups in relation to serum protein (mean± standard error)

No relevant difference was observed between Group (0.042±0.037), Group II (0.063±0.091), Group III (0.036±0.033) in comparison to the Group IV (0.044±0.037) for amylase activity in saliva. For amylase activity in serum, Group I (0.606±0.250), Group II (0.592±0.177), Group III (0.608±0.341) and Group IV (0.519±0.519) has no significant relation.

For Urine microalbumin, healthy controls were taken as the reference where the mean value was observed as Group I (47.50±20.20), Group II (102.2±23.77) and Group III (311.45±83.53) (Table 5.5)(Figure 5.5).

GROUPS	**N**	**MEAN**	**SD**	**P**	**t-VALUE**
CONTROL -NEWLY DIAGNOSED	50	47.50	20.20	0.0	16.621
CONTROL -1-2 YEARS	50	140.02	23.77	0.0	41.643
CONTROL -MORE THAN 3 YEARS	50	311.45	83.53	0.0	26.365

Table 5.5: Comparison of type 2 diabetes mellitus and control groups in relation to microalbumin (mg/dl) by Paired T- test

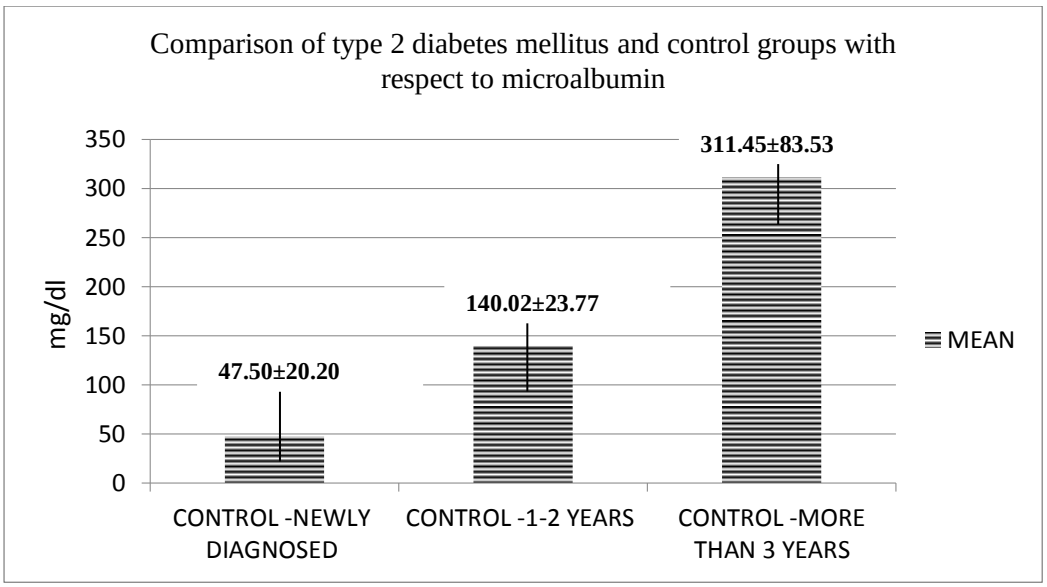


Figure 5.5: Graphical representation of Type 2 diabetes mellitus and control groups in relation to microalbumin(mean± standard error)

5.3. Study of mutations related with candidate genes and their association with type 2 diabetes in Mizo Population.

From the isolated genomic DNA, the concentration of DNA ~ 50 to 100 ng was taken for further study.

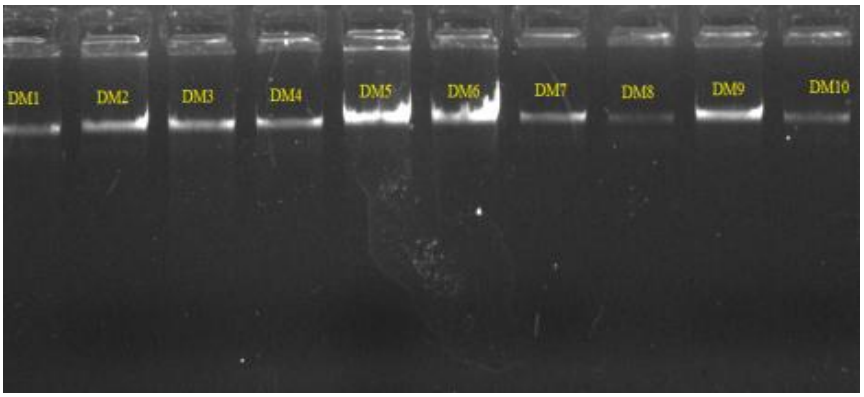


Figure 5.6: Representative Image of Genomic DNA Isolated from Blood samples = **DM1 to DM10**

5.3.1. Study of candidate genes using Sanger sequencing

Polymerase chain reactions was performed for KCNJII, PPARG, HNF1A and HNF4A genes (Figure 5.7)

Figure 5.7 A: Amplified products of KCNJ11 gene. Product size ~ 567 bp, sample no.'s - DM1 to DM8

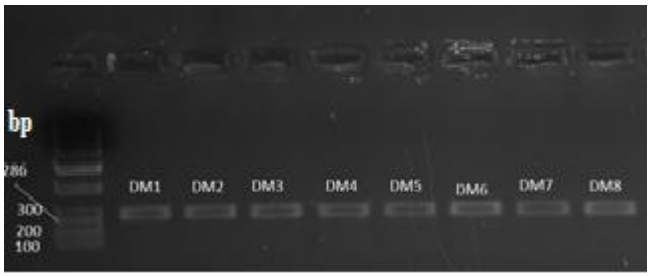


Figure 5.7 B: Amplified products of HNF1A gene. Product size ~ 286 bp, sample no.'s - DM1 to DM8

Figure 4.7 C: Amplified products of PPARG gene. Product size ~ 220 bp, sample no.'s - DM1 to DM8

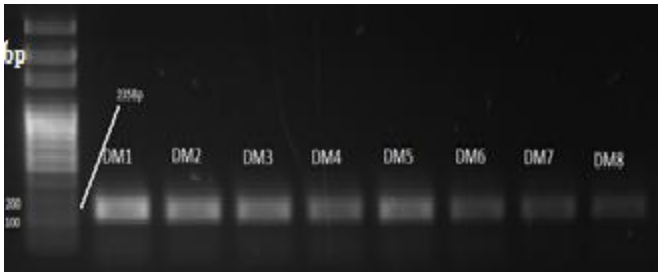


Figure 5.7 D: Amplified products of HNF4A gene. Product size ~ 235 bp, sample no.'s - DM1 to DM8

From the PCR results, 25 diabetic and 25 healthy controls were selected for Sanger sequencing and no variation were observed for KCNJII, PPARG and HNF4A genes among them. However, we observed G>C polymorphism (Figure 5.8) at chromosome 12:121438908 for HNF1A gene and it was found to be synonymous and does not have an impact on splicing.

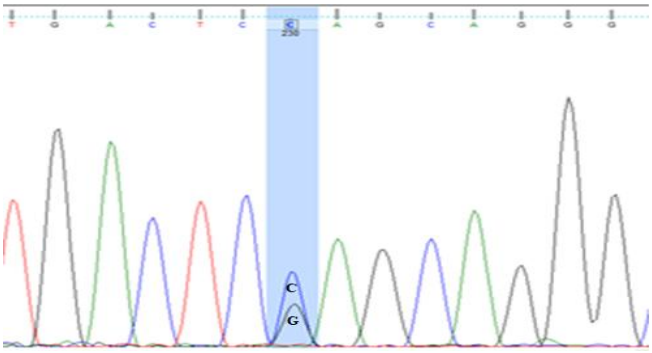


Figure 5.8: G>C polymorphism on chromosome 12:121438908 of HNF1A gene

5.3.2. Tagman Allele discrimination Assay

UCP3 variant rs1800849 and MACF1 rs2296172 were genotyped by using allele discrimination assay in 450 Type 2 diabetes cases and 437 healthy controls belonging to Mizo population.

The genotype frequencies of *UCP3* variant rs1800849 was found to be in concomitant with HWE (p-value = 0.707). The allele frequency observed for allele G in cases and controls was 0.88 and 0.874, respectively whereas the observed frequency of allele A was 0.12 in cases and 0.126 in controls. The frequency of both the alleles was almost same in cases and controls (Table 5.6). The allelic OR observed for the variant rs1800849 of *UCP3* was 0.95[0.70-1.29 at 95% CI], p-value = 0.733. We found no variation in the results even after adjustment for age, gender and BMI.

The variant rs1800849 of *UCP3* did not show any significant association with susceptibility to Type 2 diabetes in Mizo population.

For MACF1, the distribution of allele frequency was observed to be following Hardy-Weinberg's equilibrium (p=0.218). In the population group, the risk allele (G) frequency was found to be significantly higher in cases (0.13) as compared to that of controls (0.074). The variant was observed to be associated with T2D and dominant mode of inheritance (GG+AG v/s AA) with p =0.001 and OR=1.8 (1.3 - 2.8) at 95% CI, (Table 4.6) appeared most appropriate. The PAR percentage observed was 5.82% (2.22 -11.33 %) at 95% CI (Table 5.7).

Allelic Distribution			Association				
Allele	**Cases**	**Controls**	**Models**	**OR [95% CI]**	**p value**	**OR* [95% CI]**	**p value***
G	0.880	0.874	Allelic	0.95 [0.7-1.29]	0.733	-	-
A	0.120	0.126	Dominant	0.9 [0.64-1.27]	0.553	0.8 [0.6-1.20]	0.479

*Corrected with age, gender and BMI

Table 5.6 : Allelic frequency distribution and risk associated with the *UCP3* variation in Mizo population.

Gene/ SNP	**Cases (n= 450)**	**Controls (n= 437)**	**HWE**			**Allelic OR**	**p valu e**	**Dominant OR***	**p value ***
	A=0.87	A=0.926	Cases	Controls	Total population				
***MACF1*/ rs2296172**	G=0.13	G=0.074	0.272	0.98	0.218	1.87 [1.31- 2.66]	0.002	1.8 [1.3 - 2.8]	0.001

*Corrected with age, gender and BMI

Table 5.7: Allelic frequency distribution and risk associated with the *MACF1* variation in Mizoram population.

5.3.3. Agena Mass Array

A total of 768 participants, 384 with type 2 diabetes and 384 healthy controls were assessed in the study. The Mizo ethnic tribe was genotyped for 12 gene variants: LRRC31 (rs16847897), HLA-G (rs1063320), VEGFA (rs2010963, rs3025020), CD2AP (rs9369717, rs9349417), ELMO1 (rs741301, rs1882080), TERF1 (rs6982126), TCF7L2 (rs290475), FTO (rs9939609) and ND3 (rs2853826). Quality control analysis was done for the 12 variants and out of which seven (07) variants was found to pass the QC analyses and in concordance with HWE and undergo further analysis for their association with type 2 diabetes. Among the 7 variants, 6 variants were found to be significantly associated with T2D. We also compared some clinical characteristics of genotypes for each variant and some variants shows effect on the clinical characteristics associated with T2D (Table 5.8).

GENE	SNP ID	POSITION	OR	P	HWE	EA	OA
LRRC31	rs16847897	169850328	1.286	**0.051**	0.051	C	G
HLA-G	rs1063320	29830972	1.254	0.260	0.434	C	G
VEGFA	rs2010963	43770613	2.673	**0.00**	0.298	C	G
VEGFA	rs3025020	43781373	0.221	**0.00**	0.685	T	C
CD2AP	rs9369717	47586732	0.808	0.303	0.050	G	T
CD2AP	rs9349417	47612921	0.743	0.097	0.479	G	A
ELMO1	rs741301	36878390	0.339	**0.00**	0.082	C	T
ELMO1	rs1882080	36916236	0.605	**0.00**	0.122	A	G
TERF1	rs6982126	73027388	1.701	0.113	0.26	T	C
TCF7L2	rs290475	113114260	1.501	**0.027**	0.596	C	T
FTO	rs9939609	53786615	0.756	0.134	0.314	A	T

EA: Effected Allele, OA: Other Allele

Red denotes highly associated while yellow denotes less likely associated and black denotes not associated

Table 5.8: Plink data showing associated genes

5.3.4. Whole Exome sequencing

From the annotated variants of 4 diabetic and 6 controls, the variants were filtered and observed based on-

a) Prediction software results

Based on the prediction software's, only disease causing/pathogenic variants were given the most priority where seven (07) similar variants were found in all the diabetic samples; however the variants were already reported in the database (Table 5.9). The variants like 328G>C (KRT18), 997G>T (CYP4A11), 2368T>C (SLC4A3), 508G>A (SLC26A5), 1659C>T (KCNS1) and 650C>A (ABCD1) variants were not reported in the database. From the prediction scores of Mutation Taster and Polyphen 2, the unreported variants were found to be pathogenic with a score range ≥ 0.99. From the Grantham matrix score, two variants 328G>C (KRT18) and 2368T>C (SLC4A3) were found to have conservative replacement with a score range between 0-50 while the variants like 508G>A (SLC26A5), 1659C>T (KCNS1) and 650C>A (ABCD1) were found to be moderately conservative with a score range between 51-100. The variant 997G>T (CYP4A11) undergoes radical substitution with a score range >100.

Chr no	position	Ref allele	Alt allele	Gene	Amino acid change
7	G146T	G	T	PRSS1	G49V
7	T158A	T	A	PRSS1	I53N
15	G140A	C	T	CHRFAM7A	R47H
9	T172C	A	G	AQP7	Y58H
9	C501G	C	G	CACNA1B	N167K
17	G415A	G	A	KCNJ12;KCNJ18	E139K
7	G194A	C	T	PSPH	R65H

Table 5.9: Non-synonymous variants observed in all the diabetic samples

16 missense variants were found out of which 10 variants were previously reported (Table 5.10) and 6 variants were not reported (Table 5.11) (Figure 5.9). From the prediction scores of Mutation Taster and Polyphen 2, the unreported variants were found to be pathogenic with a score range ≥ 0.99. From the Grantham matrix score, two variants 328G>C (KRT18) and 2368T>C (SLC4A3) were found to have conservative replacement with a score range between 0-50 while the variants like 508G>A (SLC26A5), 1659C>T (KCNS1) and 650C>A (ABCD1) were found to be moderately conservative with a score range between 51-100. The variant 997G>T (CYP4A11) undergoes radical substitution with a score range >100.

Chr no	position	Ref allele	Alt allele	DNA Change	Gene	Amino acid change	Exonic Functon
2	215845346	G	A	G2089A	ABCA12	G697S	Missense
4	113462397	G	C	C5452G	ZGRF1	Q1818E	
5	43659300	G	A	G2089A	NNT	G697S	
6	49425591	T	C	A566G	MUT	N189S	
7	56087374	C	T	G194A	PSPH	R65H	
7	142460339	G	A	G512A	PRSS1	C171Y	
8	104412688	T	C	A899G	SLC25A32	Y300C	
9	33796746	G	T	G125T	PRSS3	G42V	
17	48755303	C	T	C3577T	ABCC3	R1193W	
22	18609481	G	A	G538A	TUBA8	G180R	

Table 5.10: Missense variants previously reported in the Databases

Gene	Mutation	Amino Acid Change	Mutation Taster/ Polyphen2 / Grantham Matrix score	Effect of the Mutations by HOPE	Reliability Index (RI) / DDG	Stability of protein
KRT18	328G>C	A92P	Disease causing/ 0.992/27	Amino acid with different property	7.0 /7.0	Increase
CYP4A11	997G>T	G322V	Disease causing/ 0.999/109	Mutant residue is bigger than the wild-type residue.	1/-1.34	Decrease
SLC4A3	2368T>C	F726L	Disease causing /1/22	Mutant residue is smaller than the wild-type residue	3 /-1.00	Decrease
SLC26A5	508G>A	G91E	Disease causing/ 0.999/98	Residue charge changes from Neutral to negative.	6/0.70	Increase
KCNS1	1659C>T	T421I	Disease causing/ 0.998/89	Residue located on the surface of the protein disturbs the interactions with other molecules	1/7.0	Increase
ABCD1	650C>A*	P84H	Disease causing/ 1/77	Hydrophobic interactions will be lost.	7 /7.0	Decrease

Table 5.11: Novel Variants not reported in the Databases

*** Indicates homozygous mutation**. DDG= DDG: DG (NewProtein)-DG(WildType) in Kcal/mol

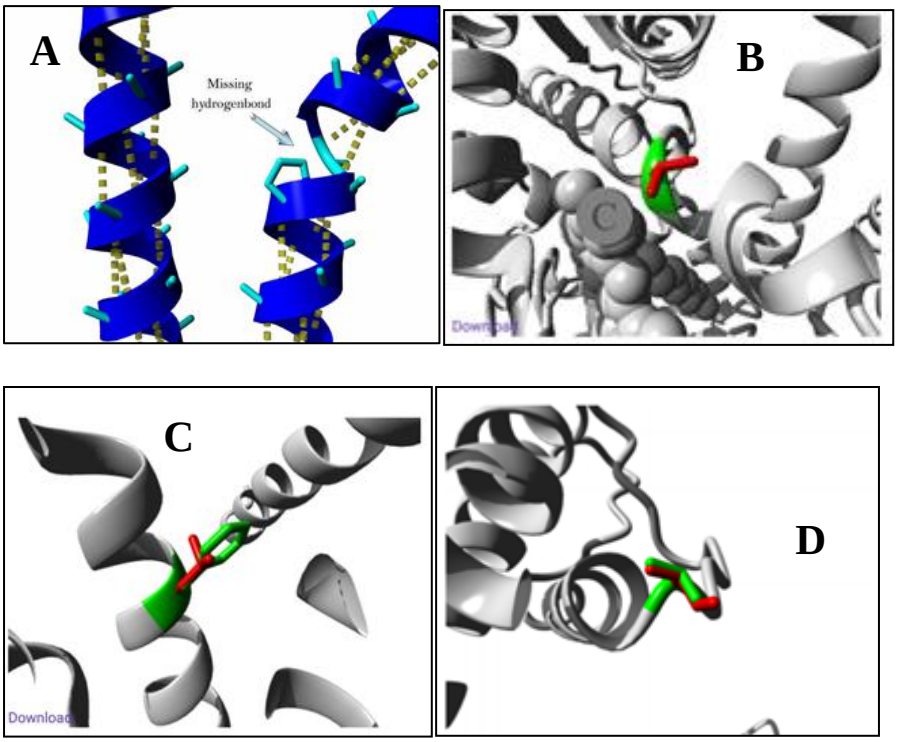


Figure 5.9: 3-D structure of the mutation in samples. Both the wild-type and mutant side chain are shown in green and red, respectively. The rest of the protein are shown in grey. No 3-D structures were available for SLC26A5 and ABCD1 gene due to a lack of structural information

A= mutation of a Alanine into Proline at position92 for KRT18 gene
B= mutation of a Glycine into Valine at position322 for CYP4A11 gene
C= mutation of a Phenylalanine into Leucine at position726 for SLC4A3 gene.
D= mutation of a Threonine into Isoleucine at position421 for KCNS1 gene.

Gene	Pathway involved (KEGG pathway)	Reported Phenotype of the genes by Genecards
KRT18	Estrogen signaling	cleft palate, cleft lip
CYP4A11	Fatty acid degradation	blood and serum metabolite measurement
SLC4A3	-	cardiovascular, respiratory, vision, aging metabolism, nervous system, behavior,
SLC26A5	-	Deafness, Neurosensory, Autosomal recessive
KCNS1	-	Mean platelet volume, schizophrenia, well-being measurement, bipolar disorder
ABCD1	Peroxisome	rheumatoid arthritis, celiac disease, mental or behavioral disorder

Table 5.12: Pathways and phenotype of the genes having variants which were not reported for Diabetes

b) Analysis based on panel genes

Based on the result of the analysis using 582 diabetic genes were taken from GWAS catalogue. We observed that-

1. Four (4) variations were found to be pathogenic using the prediction scores

Chrm no	position	Ref allele	Alt allele	gene	AA change	Exonic function
4	42509101	G	A	ATP8A1	C2018T	Non-synonymous
9	94486693	C	T	ROR2	G695R	
19	33878977	C	T	PEPD	G971A	
19	33878977	C	T	PEPD	G1040A	

Table 5.13: Reported pathogenic variants based on GWAS panel

According to mutation taster prediction, a total of 2937 variations were not reported and predicted to be polymorphisms.202 variations were found to be disease causing and among the disease causing variations [8 were found in the exonic region (Table 5.14; Table 5.15, Figure 5.10) and 142 in the intronic region.

From the Mutation taster and Polyphen2 score, the variants like 821A>T (YTHDC2), 931G>T (PINX1), 1579C>T (KCNQ1), 3280C>A (TNRC6A), 48C>A(TACO1), 6035A>T(LAMA1), 805C>A(ACP7) and 806A>G(ACP7) were observed to have a damaging effect with scores ≥ 0.99. The Grantham matrix score shows that the variants 821A>T (YTHDC2), 1579C>T (KCNQ1), 3280C>A (TNRC6A), 48C>A (TACO1) and 806A>G (ACP7) have a score between 0-50 and the replacement has been found to be conservative. The replacement from the variant 805C>A (ACP7) has been observed to be moderately conservative (score range 51-100) while the variants like 931G>T (PINX1) and 6035A>T (LAMA1) have radical substitution or replacement with a score range 101-150 and ≥150.

Gene	Mutation	Amino Acid Change	Mutation Taster/Polyphen2 /Grantham Matrix score	Effect of the Mutations by HOPE	Reliability Index (RI)/ DDG	Stability of protein
YTHDC2	821A>T	Y274F	Disease causing/ 0.999/ 22	Mutation causes an empty space in the core of the protein.	5/-0.35	Decrease
PINX1	931G>T	D311Y	Disease causing/ 0.999/ 160	Residue charge changes from negative to Neutral	3/-1.45	Decrease
KCNQ1	1579C>T	K527L	Disease causing/ 1/ 22	Mutant residue is smaller and might lead to loss of interactions.	7.0 /8	Increase
TNRC6A	3280C>A	P1094T	Disease causing/ 1/ 38	Wild-type residue is more hydrophobic than the mutant residue.	8/ -2.23	Decrease
TACO1	48C>A	C16Y	Disease causing/ 1/ 6	Mutant residue is bigger, this might lead to bumps	4 /-0.93	Decrease
LAMA1	6035A>T	K2012I	Disease causing/ 0.992/ 102	Mutant residue is more hydrophobic than the wild-type residue	1/0.30	Decrease
ACP7	805C>A	Q269K	Disease causing/ 0.999/ 53	Mutation introduces a charge at this position, this can cause repulsion between the mutant residue and neighboring residues.	4/-0.78	Decrease
ACP7	806A>G	Q269R	Disease causing/ 1/43	Mutant residue is bigger than the wild-type residue.	1/0.04	Increase

Table 5.14: Unreported pathogenic variants based on GWAS panel with HOPE prediction DDG= DDG: DG (NewProtein)-DG(WildType) in Kcal/mol

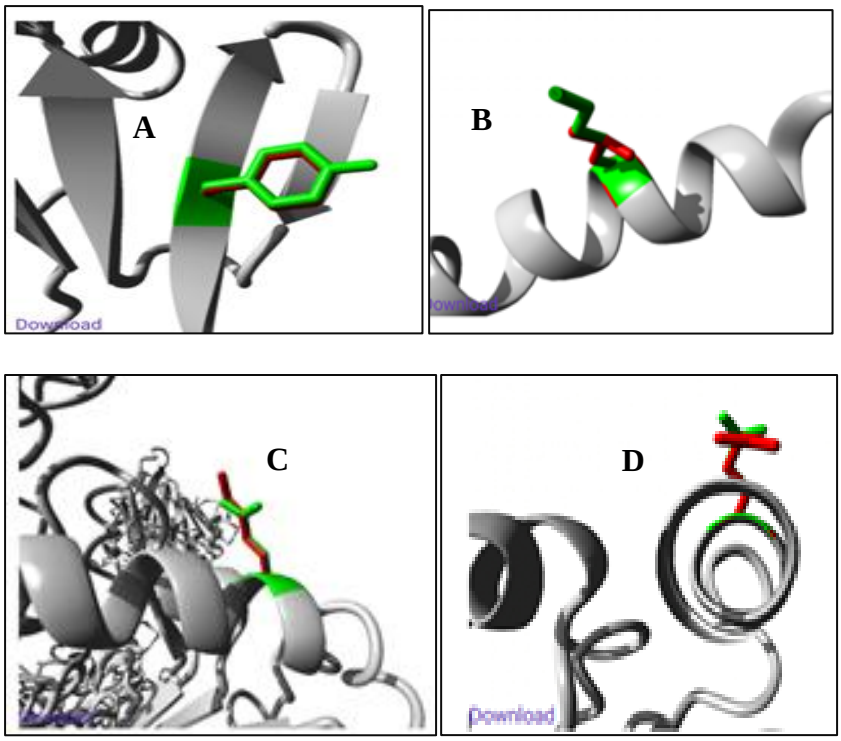


Fig 5.10:3-D structure of the mutation in samples. Both the wild-type and mutant side chain are shown in green and red, respectively. The rest of the protein is shown in grey. **No 3-D structures were available for PINX1,TRNC6A,TACO1 and LAMA1 gene due to a lack of structural information.**

A=mutation of a Tyrosine into a Phenylalanine at position 274 for **YTHDC2** gene.

B=mutation of a Lysine into a Leucine at position 527 for **KCNQ1** gene

C=mutation of a Glutamine into a Lysine at position 269 for **ACP7** gene.

D=mutation of a Glutamine into a Arginine at position 269 for **ACP7** gene.

Gene	Pathway involved (KEGG pathway)	Reported Phenotype of the genes by Gene cards
YTHDC2	-	fetal hemoglobin measurement, sickle cell anemia
PINX1	-	systolic blood pressure, alcohol drinking, pulse pressure, smoking status measurement
KCNQ1	Pancreatic secretion	type 2 diabetes mellitus,A1C measurement, BMI, glucose measurement
TNRC6A	-	intelligence, systolic bp, pulse pressure, alzheimers disease
TACO1	-	cardiovascular, vision, aging
LAMA1	Pathways in cancer	type 2diabetes, cardiac arrest, Alzheimer, smoking behavior, lung cancer
ACP7	-	type 2 diabetes, intelligence
ACP7	-	type 2 diabetes, intelligence

Table 5.15: Pathways and phenotype of the genes having variants which were reported

Among 202 intronic variations 9 InDel's were found in in the exonic region (Table 5.16) and 37 InDel's in the intronic region (Table 5.17).

Chr no	position	Ref allele	Alt allele	DNA Change	Fuctional ref.gene	gene	Amino acid change	Exonic function
1	39907835	T	-	c.18902_18902delT	exonic	MACF1	Y6302Tfs*16	frameshift deletion
7	127236017	C	-	c.399_399delC		FSCN3	L134Sfs*24	
8	96047753	-	A	c.369_370insA		NDUFAF6	I124Nfs*4	
13	24804890	G	-	c.96_96delG		FREM2	L33Cfs*9	
16	24804890	G	-	c.3272_3272delG		TNRC6A	E1093Nfs*22	
17	6945683	C	-	c.818_818delG		SLC16A11	G273Afs*46	
17	1840111	G	-	c.1005_1005delC		RTN4RL1	original stopcodon lost, results in prolonged protein	
19	19613321	C	-	c.1760_1760delC		GATAD2A	T587Sfs*35	

Table 5.16: InDel's observed in the exonic region and not reported for any disease so far.

Chr no	position	Ref allele	Alt allele	DNA Change	Fuctional ref.gene	gene
1	51412439	TA	-	g.13496_13497delTA	intronic	FAF1
1	176877269	ACACACAC	-	g.256834_256841delGTGTGTGT		ASTN1
1	44723785	C	-	g.97148_97148delG		ERI3
1	241161830	AGC	-	g.358699_358701delGCT		RGS7
1	44803699	A	-	g.17234_17234delT		ERI3
2	236476608	T	-	g.73876_73876delT		AGAP1
2	43586926	T	-	g.236260_236260delA		THADA
3	71117276	AT	-	g.515864_515865delAT		FOXP1
3	71154567	A	-	g.478574_478574delT		FOXP1
3	65506999	A	-	g.517511_517511delT		MAGI1
3	173518044	TTT	-	g.403971_403973delTTT		NLGN1
4	164910439	GAGA	-	g.394761_394764delTCTC		MARCH1
5	101594456	ACAC	-	g.37795_37798delGTGT		SLCO4C1

Cont.

Chr no	position	Ref allele	Alt allele	DNA Change	Fuctional ref.gene	gene
5	53297642	G	-	g.308771_308771delC	intronic	ARL15
5	155871968	A	-	g.574615_574615delA		SGCD
7	77786455	T	-	g.1296436_1296436delA		MAGI2
7	151329234	GGA	-	g.244977_244979delTCC		PRKAG2
7	157636341	C	-	g.744140_744140delG		PTPRN2
7	103527039	TTT	-	g.403971_403973delTTT		NLGN1
7	173518044	GAGA	-	g.394761_394764delTCTC		THADA
7	103527039	A	-	g.102925_102925delT		RELN
7	18631101	TG	-	g.504530_504531delTG		HDAC9
7	103367644	T	-	g.262320_262320delA		RELN
7	43410270	A	-	g.258073_258073delA		HECW1
8	105601192	-	CGCCGACGCCGC	cDNA.60_61insGCGGCGTCGGCG g.225_226insGCGGCGTCGGCG		LRP12
8	103527039	A	-	g.102925_102925delT		RELN

Cont…

Chr no	position	Ref allele	Alt allele	DNA Change	Fuctional ref.gene	gene
8	96047112	TT	-	g.139118_139119delTT	intronic	NDUFAF6
8	10075237	TC	-	g.163460_163461delTC		MSRA
9	8674754	A	-	g.1937970_1937970delT		PTPRD
9	27997355	A	-	g.672930_672930delA		LINGO2
9	3977392	A	-	g.371001_371001delT		GLIS3
9	9005680	A	-	g.1607044_1607044delT		PTPRD
10	121044221	G	-	g.77121_77121delG		GRK5
10	101969849	TTT	-	g.19526_19528delAAA		CHUK
11	72424215	C	-	g.80430_80430delG		ARAP1
11	128649161	T	-	g.94116_94116delT		FLI1
16	69602278	TGTA	-	g.3282_3285delTGTA		NFAT5
16	24741079	G	-	cDNA.64_64delG g.64_64delG		TNRC6A
16	53778232	T	-	g.40358_40358delT		FTO

Table 5.17:InDels observed in intronic region and has not been reported in the database.

From the results of unreported pathogenic variants from both the prediction software's as well as the panel genes, HOPE prediction software reports that the mutations have an effect on the protein structures. On further prediction of the protein stability using i mutant 2.0, the result reports that the change in the amino acid of the genes like CYP4A11 (G322V), SLC4A3 (F726L), ABCD1(650C>A), YTHDC2 (Y274F), PINX1 (D311Y), TNRC6A (P1094T), TACO1 (C16Y), LAMA1 (K2012I) and ACP7 (Q269K) decreases the stability of the protein with DDG value <0. The list of novel genes identified in our study and reported as possible prognostic or diagnostic markers for various diseases are given in Table 5.18.

Gene	Prognostic Marker	Diagnostic marker
KRT18	Gastrointestinal carcinomas, Cancers of Breast, Prostrate, Cervical and Liver	-
CYP4A11	Hepatocellular carcinoma	Coronary artery disease
SLC4A3	Breast cancer	-
KCNS1	Pain risk in human	Multiple chronic pain states
ABCD1	X-linked adrenoleukodystrophy	X-linked adrenoleukodystrophy
YTHDC2	Colorectal cancer	colorectal cancer
PINX1	Breast cancer	-
KCNQ1	Hepatocellular carcinoma, colon and ovarian	Jervell and Lange-Nielsen syndrome
TNRC6A	Human clear cell renal cell carcinoma, gastric and colorectal cancers	-
TACO1	Pancreatic cancer, leigh syndrome	-
LAMA1	-	cerebellar malformations, Joubert syndrome

Table 5.18: List of genes from our study previously reported as possible prognostic or diagnostic markers

5.3.5. Whole mitochondrial sequencing

From the 28 diabetic DNA samples and 12 controls that were taken for sequencing and analysis. It has been observed that

1. Ten (10) common variants were found both in cases and controls out of which 2 variants were previously reported for diabetes (Table 5.19).

cDNA change	Ref allele	Alt allele	gene	Aa change	GB Freq (%)	Conservation (%)	Information
A334G	A	G	ATP6	T112A	98.64%	71.11%	Reported
A340G	A	G	ND3	T114A	33.91%	42.22%	Not reported
C20T	C	T	CYTB	T7I	76.96	48.89	
A580G	A	G	CYTB	T194A	98.68	17.78	
A175G	A	G	ATP6	T59A	33.99	42.12	
G1372A	G	A	ND5	A458T	7.17	33.33	Reported
G991A	G	A	ND2	A331T	0.09	17.78	Not reported
T140C	T	C	ND4L	M47T	0.00	37.78	
G70A	G	A	ND5	V24I	2.62	6.67	
G1592C	G	C	ND5	S531T	4.84	11.11	

Table 5.19: Common variants of Cases and Control with MITOMASTER Scores

GB frequency= Giordano Bruno frequency

2. Eighteen (18) case specific variants were observed out of which 3 variants were previously reported for diabetes (Table 5.20).

cDNA change	Ref allele	Alt allele	gene	Aa change	GB Freq (%)	Conservation (%)	Information
A364G	A	G	ND2	T122A	0.93	17.78	Reported
T926C	T	C	ND1	I309T	1.68	80	Not reported
G22A	G	A	ND2	V8I	1.82	4.44	Reported
G271A	G	A	COX3	V91I	0.00	91.11	Not reported
C709A	C	A	ND2	L237M	4.78	22.22	Reported
C49T	C	T	ATP8	L17F	3.98	31.11	Not reported
A817G	A	G	ND5	I273V	0.04	51.11	
G1138A	G	A	CYTB	A380T	0.03	4.44	
G58A	G	A	ATP6	A20T	5.21	17.70	
T1774C	T	C	ND5	F592L	0.96	11.11	
A355G	A	G	ND2	T119A	2.94	82.22	
A133G	A	G	ATP6	T45A	0.22	17.78	
C268T	C	T	ATP6	H90Y	2.83	73.33	
G184A	G	A	ND4L	A62T	0.00	2.22	
A769G	A	G	ND5	I257V	7.60	44.44	
A787G	A	G	ND1	T263A	0.31	71.11	
A717T	A	T	ND2	W239C	0.08	17.78	

Table 5.20: Sample specific variants with MITOMASTER scores

For the variations present commonly between the cases and controls, the replacement were mostly found in the hydrophobic regions of the mitochondrial proteins (90% of the variant) with hydrophobicity scale ranging from 0.37 to 1.26. For synonymous variants, the replacements were mostly found in the hydrophobic regions (60%) and hydrophilic regions (40%) of the mitochondrial proteins.

The ratio of Non-synonymous to Synonymous mutations (NS/S) in common variants and case specific are 10:18 and 9:4. The variants observed in the diabetic

cases only belonged to mitochondrial complex I, IV and V while the variants common in both the cases belonged to complex I, III, IV and V. In majority, the non-synonymous substitutions in case specific as well as common groups decreased the stability of the protein function (Table 5.21). The dN/dS ratio of germ line mutations were found to be greater than 1 for complexes I and IV (15.87 and 3.76) which indicates a positive selection while that of complexes III and IV (0.264 and 0.686) were found to be stabilizing the selection for non-synonymous germline mutations in mtDNA in diabetic-healthy individuals.

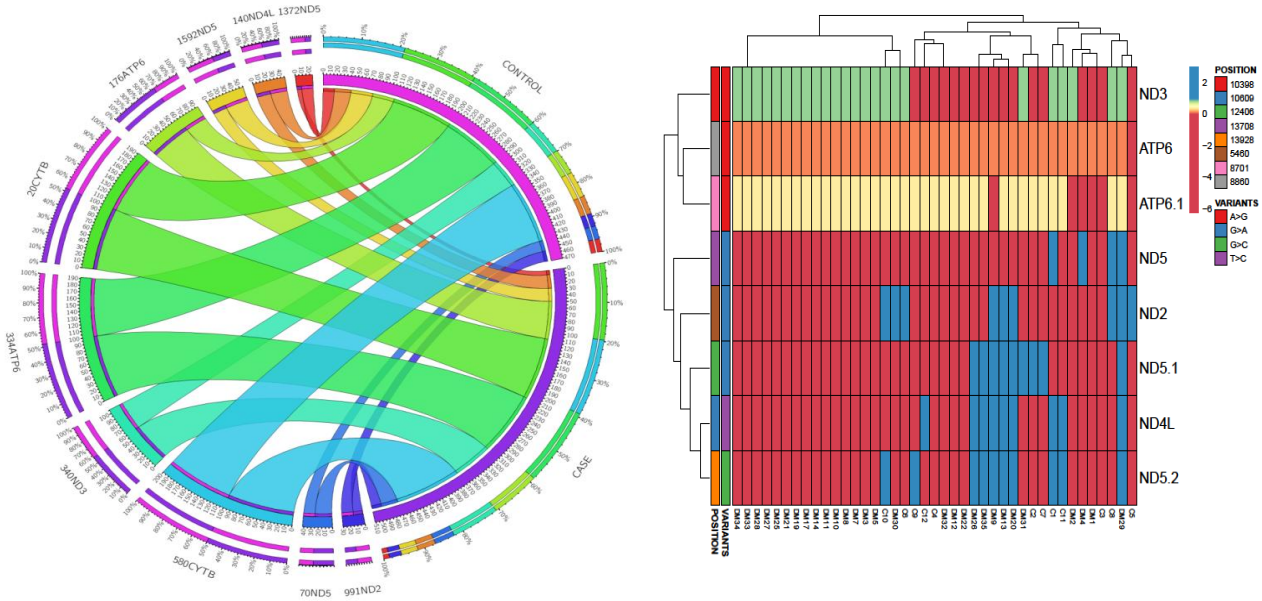


Figure 5.11: Circos plot and heatmap representing Common variants of Cases and Control

Diabetic specific variants			
Substitution	**Protein change**	**Δ Δ G**	**Effect on protein**
8584G>A	ATP6:A20T	-1.835	Decrease stability
8659A>G	ATP6:T45A	-0.846	
8794C>T	ATP6:H90Y	-0.036	
8414C>T	ATP8:L17F	-1.468	
4232T>C	ND1:I309T	-1.998	
4093A>G	NDI:T263A	-0.527	
4833A>G	ND2:T122A	-0.545	
4491G>A	ND2:V8I	0.209	Increase stability
5178C>A	ND2:L237M	-0.902	Decrease stability
4824A>G	ND2:T119A	-0.399	
5186A>T	ND2:W239C	-0.841	
10653G>A	ND4L:A62T	-1.187	
13153A>G	ND5:I273V	-0.948	
14110T>C	ND5:F592L	-1.304	
13105A>G	ND5:I257V	-0.660	
9477G>A	COX3:V91I	-0.680	
15884G>A	CYTB:A380T	-1.012	
Common variants			
A8860G	ATP6:T112A	-0.398	Decrease stability
A8701G	ATP6:T59A	-0.785	
G5460A	ND2:A331T	-1.247	
A10398G	ND3:T114A	-0.928	
T10609C	ND4L:M47T	-1.276	
G13708A	ND5:A458T	-0.948	
G12406A	ND5:V25I	-0.192	
G13928C	ND5:S531T	-0.693	
C14766T	CYTB:T7I	-0.146	
A15326G	CYTB:T194A	-0.427	

Table 5.21: Potential impact of non-synonymous case specific as well as common variants as predicted by Pmut.

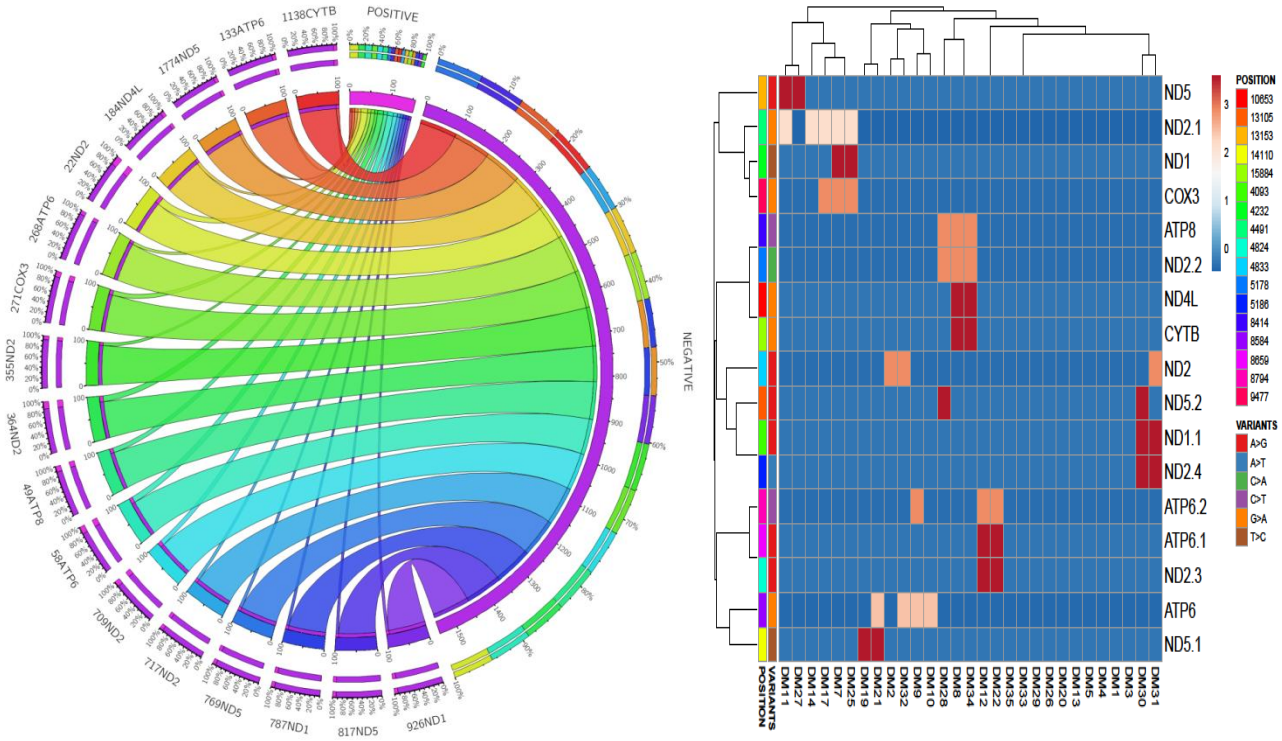


Figure 5.12: Circos plot and heat map representing Case specific variants

The mitochondrial D-Loop region was analyzed where ten (10) common variants for both cases and controls were found in upstream of HV-1 and HV-2 (Table 4.20). For both the groups (case specific and common), the variation has been mostly observed in tRNA-Phe region of mtDNA. The variation has also been observed in other regions like 12S rRNA, 16S rRNA, tRNA-Ala, tRNA-Asp and tRNA- Thr.

Gene	position	ref	alt	information
RNR1	73	A	G	reported, not for diabetes
	152	T	C	
	247	G	-A	
CYTB	263	A	G	reported for diabetes
	302	A	AC	reported, not for diabetes
	16189	T	C	reported for diabetes
	16223	C	T	
	16304	T	C	reported ,not for diabetes
	16311	T	C	
	16362	T	C	

Table 5.22: Common variants of Cases and Control in D-LOOP

Eighteen (18) sample specific variants were found in upstream of HV-1 and HV-2

Gene	position	ref	alt	information
RNR1	56	A	AC	reported, not for diabetes
	57	T	C	
	94	G	-A	
	143	G	A	
	146	T	C	
	151	C	T	
	173	T	C	
	183	A	G	
	195	T	C	
	200	A	G	
	204	T	C	reported for diabetes
	207	G	A	reported, not for diabetes
	234	A	G	
	235	A	G	
CYTB	16093	T	C	reported for diabetes
	16136	T	C	reported, not for diabetes
	16319	G	A	
	16381	T	C	

Table 5.23:Sample specific variants in D-LOOP

Chapter VI

Discussion

6. Discussion

The objective of this study focuses on the role of demographic factors, biochemical profiles as well as genetic factors which may contribute to the onset of diabetes in Mizo population. Although the prevalence of T2D relates with family history of diabetes, obesity as well as age dependent, there are also involvement of certain genetic factors contributing from different parts of the genome (Karen et al., 2003). Firstly from the study, the unique lifestyle and dietary habits of Mizo population leads to the attribution towards the prevalence of type 2 diabetes in Mizoram. The risk of type 2 diabetes was observed to be higher in patients consuming an animal fat product prepared semi-dry in bottle gourd (*Lagenaria siceraria*) known as 'Saum' which is also a derivative of pork fat. Although it exhibits distinct astringency it has no significant organoleptic qualities. Saum is one of the traditional foods for the local people although it exemplifies adverse health characteristics due to the presence of high saturated fat/cholesterol content (Hooper et al., 2001; Surajit et al., 2018).

Smoked meat and excess salt consumption may act as secondary risk factor for type 2 diabetes in Mizoram. The people of Mizoram still practice a variety of food processing habits that were passed on from their fore-fathers where only the smoke and the heat directly affects the meat (generally not barbecued) in the traditional smoking process. High-heat cooking many produce harmful chemicals such as polycyclic aromatic hydrocarbons, heterocyclic aromatic amines, and nitrosamines (from nitrates and nitrites added to meats as a preservative) which may lead to an inflammatory response, interfere with the normal production of insulin. High salt or excess salt intake may lead to insulin resistivity which in turn may lead

to hyperglycemia. Betel nut chewing and tuibur consumption also contribute to the risk of developing type 2 diabetes. Studies shows that betel nut chewers may develop type 2 diabetes at an early stage which may peaked later in life around the age of 60-69 years (Chin et al., 2010).

Tuibur is a tobacco infused water and since tobacco is known to be linked with many diseases and nicotine's ability to affect certain antioxidant enzymes like lipid peroxidase, superoxide dismutase, etc., may in turn contribute to the development of type 2 diabetes (Cooper et al., 2006). In the present study, tuibur acts a confounding factor for type 2 diabetes development. Studies from different population reported that smoking has an impact in T2DM development (Cooper et al., 2006), as chronic smokers have a higher risk of insulin resistivity (Xie et al.,2009; Facchini et al., 1992; Carlsson et al.,2004; Eliasson et al.,2003). However in our study, we observed no significant relationship between smoking and T2DM in this population. The levels of Creatinine may also play a crucial role in depicting the risk for type 2 diabetes as the fluctuation in the level can arise due to cigarette smoking, individuals with larger body weight and height, male gender and individuals with older age, which in turn may have a regulatory effect with glucose uptake, thus leading to insulin resistance (Bamanika et al.,2016; Takeuchi et al.,2018).

Secondly, for the study involving saliva and serum, the diabetic participants were classified into three groups ie., newly diagnosed (diabetic drug naïve),1-2 years diagnosed and more than three years diagnosed. This classification is done with the assumption of whether the prescribed diabetic drug have any effect on the protein and amylase levels in the saliva as well as serum along with the duration of diabetes.

For both the saliva and serum, a fluctuation in the total protein concentration levels was observed in the present study. In general, it has been reported that the variation in protein concentration can occur due to any of following three changes:- in the rate of their anabolism, rate of their catabolism, and in the volume of distribution (Marshall et al.,2004). Each protein is known to have a distinctive half-life when it is circulated and in certain diseases, in this case like type 2 diabetes mellitus, the half-life of the protein is known to be altered (Murry et al., 2012). This matter can be added to the observed fluctuation in protein concentration. The slight increase in protein concentration for patients with 1 to 2 years diagnosed can be explained that due to medications consumed by diabetes patients, pharmacokinetics of the metabolic pathways involved in protein metabolism might be controlled (Reinke et al.,1987; Watkins et al.,1988) . However, since type 2 diabetes mellitus can influence the plasma protein-binding of some drugs, which may be through glycosylation or displacement of proteins (Gwilt et al.,1991). This may explain the decrease in protein concentration when the patient is diagnosed for more than 3 years. It has also been observed that the levels of microalbumin increases with the increase in duration of diabetes. Several studies have reported that the prevalence of microalbuminuria can reflect the major chances of ethnic population groups and their susceptibility to diabetic nephropathy (Vijay et al.,1993; Parving et al., 2001).

Thirdly, the study involving candidate genes was done using Sanger sequencing of certain genes like KCNJ11, PPARG, HNF1A and HNF4A. No variations was observed in the KCNJ11, PPARG and HNF4A genes which may be due to the contribution of ethnicity in the disease development (Abate et al., 2003). Synonymous polymorphism has been observed on chromosome 12 of HNF1A gene

for all the 25 cases and healthy controls. This can be explained that before the onset of type 2 diabetes individuals with HNF1A mutations has dysfunction in the pancreas and patients as well as healthy individuals with HNF1A mutant carriers may have glycosuria (Stride et al., 2005).

In the second approach, genotyping of UCP3 variant rs1800849 and MACF1 gene variant rs2296172 was carried out for 450 type 2 diabetes cases and 437 healthy controls. The UCP3 gene has been reported to have protective role against insulin resistance (Krook et al., 1998). The variant was first reported to be associated with Type 2 diabetes in French population (Meirhaeghe et al., 2000). The variant has been reported to have strong association with type 2 diabetes (Xu et al., 2011; de Souza et al.,2013). This replication study has been performed in order to find out the association of UCP3 gene variant rs1800849 (G>A) and we found no significant association of the variant with T2D in the Mizo population in contrast to earlier study in another population group of India (Radha et al. 2011). However, these findings are in agreement with the genetic association studies in European and Danish population where no significant association of the variantrs1800849 of UCP3 with T2D was observed (de Souza et al., 2013; Dalgaard et al., 2001). In the same manner, MACF1 gene variant rs2296172 has been replicated and has been observed that the variant may act as a risk variant in the development of type 2 diabetes. In European population (Albrechtsen et al., 2013), the G allele frequency in individuals with type 2 diabetes was 0.23 whereas, in an earlier study from India in Bania population from Punjab (Sharma et al., 2017) it was 0. 17. Further, in the present population the G allele frequency was observed to be 0.13 in T2D cases. This could be attributed to

genetic heterogeneity and presence of variant and gradient of risk allele frequency in Indian populations (Sharma et al., 2017).

In the third approach, Agena mass array was carried out for 384 type 2 diabetes cases and 384 healthy controls and 12 gene variants were genotyped. QC analyses and HWE has been carried out for which 6 variants were found to be significantly associated with T2D (OR=0.221-2.673, p value ≤ 0.0). The associated variants are discussed as follows:

rs2010963 and rs3025020 variants of *VEGFA*

The variant rs2010963 is located at 5' UTR and rs3025020 located at intronic position ofchromosome 6 of the vascular endothelial growth factor A (*VEGFA*) gene. The allele frequency distribution of the variants rs2010963 and rs3025020 in the population followed the HWE 0.298 and 0.685, respectively. It has been observed that the variant rs2010963 variant rs3025020 (T/C) were observed to be significantly associated with T2D. The observed OR for the variant rs2010963 and rs3025020 was 2.3 and 0.147, respectively and p-value 0.0 after correction for age, gender and BMI

rs741301 and rs1882080 variants of *ELMO1*

The variant rs741301and rs1882080 are located at the intronic position of chromosome 7 of the Engulfment and Cell Motility 1 (*ELMO1*) gene. The allele frequency distribution of the variants rs741301 and rs1882080 of *ELMO1* in the population followed the HWE 0.082 and 0.122, respectively. The variant rs741301(C/T) and the variant rs1882080 (A/G) were observed to be significantly associated with T2D. The observed OR for the variants rs741301 and rs1882080 of

ELMO1 in Mizo Population was 0.24 and 0.5, respectively with p-values 0.3 and 0.01, respectively after correction for age, gender and BMI.

rs9369717 variant of *CD2AP*

The variant rs9369717 is located at the intronic position of chromosome 6 of CD2 associated Protein (CD2AP) gene. The allele frequency distribution of the variant rs9369717 of *CD2AP* in the population followed the HWE 0.06. We observed that the variant rs9369717 (T/G) was observed to be significantly associated with T2D. The observed OR for the variants rs9369717 of *CD2AP* in Mizo Population was 0.56 and p value 0.02 after correction for age, gender and BMI.

rs9939609 variant of *FTO*

The variant rs9939609 is located at the intronic position of chromosome 16 ofFat mass and obesity-associated protein also known as alpha-ketoglutarate-dependent dioxygenase (FTO)gene. The allele frequency distribution of the variant rs9939609 of *FTO* in the population followed the HWE 0.384. The variant rs9939609 of *FTO* was observed to be significantly associated with T2D. The observed OR for the variant rs9939609 of *FTO* in Mizo Population was 0.49 and p value 0.004 after correction for age, gender and BMI.

From whole exome sequencing, it was observed that several variants have been found to be pathogenic like the gene variants from KRT18, CP4A11, SLC4A3, SLC26A5, KCNS1 and ABCD1 and certain change in their protein structure has been predicted. Though these variations in these genes have not been reported for a particular disease, even type 2 diabetes, it has been predicted that due to the unwanted mutation in the amino acid sequence the proteins are not able to achieve

the native state and may in turn lead to a disease (Berke et al., 2003). From 582 reported diabetic genes taken from the GWAS catalogue, 202 variations were found to be disease causing and among the disease causing variations, 142 were found in the intronic region and 8 in the exonic region. Several new gene variants have also been observed for the Mizo population like YTHDC2, PINX1, KCNQ1, TNRC6A, TACO1, LAMA1 and ACP7. These variants were observed in all the diabetic samples.

In Mizo population, (A/T) change has been observed at position 274 which causes an amino acid change Tyrosine (Y) to Phenylalanine (F) for YTHDC2. The mutation is located within a domain of Helicase ATP-binding and introduces an amino acid with different properties, which can disturb this domain and abolish its function. The function of the genes includes nucleic acid binding and helicase activity. Variations in PINXI gene has been found to be associated with type 2 diabetes in Swiss German Amish community (Wessel et al., 2018). Alteration in PINX gene can lead to differences in the lipid profile of an individual (Zhang et al.,2017) and abnormalities in the lipid levels can lead to insulin resistance (Krauss et al.,2004) which may further lead to type 2 diabetes. In this study, (G/T) change has been observed at position 311 and the amino acid changes from Aspartate (D) to Tyrosine (Y). The mutation is located within a stretch of residues on the Telomerase inhibitory domain and the wild-type residue charge was negative while the mutant residue charge is neutral. Potassium voltage-gated channel subfamily Q member 1 or KCNQ1 has been reported as a susceptibility gene for type 2 diabetes in Japan population (Kasuga et al., 2011) and the gene variants like rs2237892, rs2237895,

and rs231362 has been reported to cause the risk for type 2 diabetes and impaired renal function in the Spanish population (Riobello et al., 2016). For this study, we observed (C/T) change at position 527 which leads to the change in Lysine (K) to Leusine (L). The mutant residue is more hydrophobic than the wild-type residue and the mutant residue is smaller, this might lead to loss of interactions. TNRC6A gene encodes a member of the trinucleotide repeat containing 6 protein family and its functions include post-transcriptional gene silencing through the RNA interference (RNAi) and microRNA pathways. It has been reported that alteration in the gene can delay the epithelial wound healing in diabetic individuals (Bettahi et al., 2013). C>A change has been observed in this study, where Proline changes to Threonine at position 1094. The mutation is located within the region which is sufficient for interaction with AGO1 and AGO4 genes. The differences in amino acid properties can disturb this region and disturb its function and the wild-type residue is more hydrophobic than the mutant residue and due to the hydrophobic interactions. TACO1 or Translational Activator of Cytochrome C Oxidase I is a protein coding gene and it is one of the genes regulated by overexpression of fat mass and obesity-associated gene (FTO) whose overexpression may hamper insulin signaling (Bravard et al., 2011). The change of amino acid from Cysteine (C) to alanine (A) has been observed at position 16. It has also been observed that the mutant residue is bigger which might lead to bumps and hydrophobic interactions. The variant rs8090011 of LAMA1 gene has been reported to be associated with body mass index (BMI) of type 2 diabetic individuals (Perry et al., 2012) and the gene has also been reported to have an interaction with the environment to cause the disease (Franks et al., 2013). A>T change has been observed in our study where Lysine (K) changes to

Isoleusine (I) at position 2012. The mutation is located within a domain of Laminin EGF-like 7 and the mutation introduces an amino acid with different properties, which can disturb this domain and abolish its function. It has also been observed in this study that, C>A and A>G change for ACP7 gene may lead to the change in amino acid Glutamine (Q) to Lysine (K) and Glutamine (Q) to Arginine (R) which may lead to disturb in the interactions with other molecules or other parts of the protein. The mutation introduces a charge at this position; this can cause repulsion between the mutant residue and neighboring residues. However, further functional validation is needed to prove the association of the variants with type 2 diabetes in this population.

Mitochondrial DNA sequencing has been performed from blood DNA to find out the pattern of germline mutations. Deleterious gene variants have been identified using various prediction tools in diabetic individuals which may give an insight into its susceptibility towards type 2 diabetes. In the coding and non-coding regions of the mitochondrial genome, we observed variants which may take part in determining the risk of having the disease in healthy individuals. The true incidence of defects in mitochondrial DNA in diabetes is unknown and still the topic of debate and due to the complexity of mitochondrial genetics identification of patients becomes one of the major limiting factor (Lynn et al.,1998). In our previous study, we observed a total of 6 ATPase sequence variations [ATP6 (8584 G>A, 8602 T>C, 8616 G>A, 8701 A>G, 8790 G>A), ATP 8 (8414 C>T)] and 6 ND1 (3316 G>A, 3394 T>C, 3552 T>A, 3970 C>T, 4065 A>G 4149 C>T) sequence variations at distinct nucleotide positions (Lalrohlui et al., 2016). In the present study, common mutations

which has not been reported for type 2 diabetes in cases and controls were observed, which is predicted to be disease causing by the prediction softwares for MT-ND2 (A331T),MT-ND3(T114A), MT-ND4L(M47T), MT-ND5 (V24I, S531T), CYTB (T7I, T194A) and MT-ATP6(T59A). It may be hypothesized that since mitochondria is maternally inherited, the control samples that contain the variants are at risk of having the disease and may eventually lead to the development of the disease.

There are several mtDNA mutations recognized to consistently express a phenotype which includes diabetes (Choo-Kang et al., 2002). Although mtDNA point mutations have been described it has been observed to appear much rarer to be associated with diabetes (Lynn et al., 1998). The most common mitochondrial mutation reported for type 2 diabetes is 3243A > G mutation (Suzuki et al., 2003; Murphy et al., 2008). However, this mutation has not been observed for Mizo population. Mitochondrial DNA deletions have been reported to cause diabetes in individuals with CPEO and Kearns Sayre Syndrome (KSS) (Whittaker et al., 2007).
In our study, we observed several unreported variants which are observed in the diabetic individuals. The percentage of conservation in comparison species of these variants ranges from 11 to 80% % with respect to the reference allele; this may show strength of pathogenicity of the mutant or effected allele in causing type 2 diabetes. T16189C polymorphism in the mitochondrial DNA D-Loop region has been observed which has been reported to be associated with type 2 diabetes as it has been known to increase the risk as well as leads to insulin resistivity in Asians (Park et al., 2008). Apart from this, we also observed unreported variants in this region which might not be directly associated with the disease but may account to the development and complexity of the disease. However, further structural and functional validation

is required to fully understand the pathogenicity of the variants in causing or contributing to type 2 diabetes in Mizo population.

Nevertheless, this study can act as a baseline study for the population as a numerous reports of pathogenic mtDNA mutations associated with a diabetic phenotype highlight that the pancreas is particularly susceptible to mitochondrial dysfunction .In combination with other contributing factors this study can be used as a useful pointer towards mitochondrial disease as a diagnosis.

Chapter VII

Summary and Conclusion

7. Summary and Conclusion

Type 2 diabetes has become a common global disease where the prevalence of the disease has been found to be increasing over time. The complexity of the disease makes it harder for researchers to come to an end point about the development of the disease. Several factors have been known to play a role in the development of the disease like demography, environment, genetic etc. Ethnicities being one of the factors along with dietary and lifestyle factors and since several genes play an active role in contributing to the susceptibility as well as in the development of the disease. The present study on the Genetic variation studies in candidate genes related to Type 2 Diabetes in Mizo population, Mizoram has been proposed and designed which aims to study the demographic factors and clinico-biochemical profiles and to study mutations in candidate genes and their association with type 2 diabetes in Mizo Population. The work done in this present study along with the findings is summarized as below:

- The risk of type 2 diabetes was higher in patients who consumed Sa-um which is a fermented pork fat (Odd ratio, OR:18.98,95%) .
- Smoked meat, excess salt, tuibur and paan consumption acts as potential confounding factors for type 2 diabetes in Mizo Population.
- G>C polymorphism was observed on chromosome 12:121438908 for HNF1A gene, however the polymorphism was found to be synonymous.
- No association has been observed for variant rs1800849 of UCP3 with Mizo population.
- Variants of LRRC31, HLA-G, VEGFA, CD2AP, ELMO1, TERF1, TCF7L2, FTO, ND3 are found to be associated with Mizo population.

- From whole exome analysis, 202 variations were unreported and found to be disease causing and among the disease causing variations 142 were found in the intronic region and 10 in the exonic region.
- The change in the amino acid of the genes like CYP4A11 (G322V), SLC4A3 (F726L), ABCD1 (650C>A), YTHDC2 (Y274F), PINX1 (D311Y), TNRC6A (P1094T), TACO1 (C16Y), LAMA1 (K2012I) and ACP7 (Q269K) decreases the stability of the protein with DDG value <0.
- Novel mutations from some of genes from this study like KRT18, CYP4AII, SLC4A3, KCNS1, ABCD1, YTHDC2, PINX1, KCNQ1, TNRC6A and TACO1 have been reported for prognostic markers while the genes like CYP4A11, KCNS1, ABCD1, YTHDC2, KCNQ1 and LAMA1 as diagnostic markers for various diseases including cancer.
- From whole mitochondrial genome analysis, seventeen (17) case specific variants were observed in the genes like ATP6 (8584G>A, 8659A>G, 8794C>T), ATP8 (8414C>T), ND1 (4232T>C, 4093A>G), ND2 (4833A>G, 4491G>A, 5178C>A, 4824A>G, 5186A>T), ND4L (10653G>A), ND5 (13153A>G, 14110T>C, 13105A>G), COX3 (9477G>A) and CYTB (15884G>A) out of which 3 variants were previously reported for diabetes.
- The mtSNP 8584 G>A (ATP6: A20T) was detected in 14.28% of the diabetic patients and none in the control groups. A novel Frame-shift substitution ND5: 81_81ins A at position 12417 was observed in 53.57% of diabetic individuals.

- The variants observed only in the diabetic cases belongs to mitochondrial complex I, IV and V while the variants common in both the cases belongs to complex I, III, IV and V.
- The dN/dS ratio of germ line mutations were found to be greater than 1 for complexes I and IV (15.87 and 3.76) a positive selection while that of complexes III and IV (0.264 and 0.686) were found to be stabilizing the selection for nonsynonymous germline mutations.
- Ten (10) common variants for both cases and controls were found in upstream of the D-loop gene HV-1 and HV-2 regions. Eighteen (18) sample specific variants were found in upstream of HV-1 and HV-2.
- Majority of the variants lie in tRNA-Phe in the non-protein coding region of mtDNA for both diabetic cases and common cases.
- Significant difference in the proteins of saliva, Group I (0.043), Group II (0.058), Group III (0.022) and control group IV (0.058) and serum, Group I (0.089), Group II (0.106),Group III (0.076) and control group IV (0.117) were observed and Microalbumin levels tend to increase with the duration of diabetes.

To the best of our knowledge the study involving type 2 diabetes in relation to demographic and genetic factors are reported for the first time in the present investigation which involves the Mizo ethnic tribe. Since so many modifiable as well as non-modifiable factors contribute for the development of the disease, this study can act as a baseline in providing an insight into the prevalence in the population. Thus, from this study we conclude that although unique lifestyle and dietary habits

may act as a major risk for the incidence of the disease. The role of genes in the disease development needs to be further validated to fully understand its contribution towards the incidence as well as disease development for this population.

References

Abdulfatai BO, Olusegun AO, Lateefat BO (2012) Type 2 Diabetes Mellitus: A Review of Current Trends. Oman Medical J 27 (4): 269-273.

Abdulghani H, Maria I, Constantino LM, D'Souza M, Limacher-Gisler F, LuoC,et al (2016) An Inverse Relationship Between Age of Type 2 Diabetes Onset and Complication Risk and Mortality: The Impact of Youth-Onset Type 2 Diabetes. Diabetes Care 39:823-829.

Abujbara MA, Liswi M I, El-Khateeb MS, Flanagan SE, Ellard S, Ajlouni KM (2014) Permanent neonatal diabetes mellitus in Jordan. J Pediatr Endocr Met 27 (9-10): 879–883.

Alan P (2010) Evidence-based health-care and qualitative research. J Res Nurs. 15(6) 489–493.

Albrechtsen A, Grarup N, Li Y, Sparso T, Tian G, Cao H et al (2013) Exome sequencing-driven discovery of coding polymorphisms associated with common metabolic phenotypes. Diabetologia 56(2):298-310.

Ali O (2013) Genetics of type 2 diabetes. World J Diabetes 4(4): 114-123

Almgren P, Lehtovirta M, Isomaa B, Sarelin L, Taskinen MR, Lyssenko V, Tuomi T, Groop L (2011) Heritability and familiality of type 2 diabetes and related quantitative traits in the Botnia Study. Diabetologia 54: 2811-2819.

Altshuler D, Daly MJ, Lander ES (2008).Genetic mapping in human disease. Science 322: 881–888.

American Diabetes Association (2013) Diagnosis and classification of diabetes mellitus. Diabetes Care 36(1).

American Diabetes Association (2012) Diagnosis and classification of diabetes mellitus Diabetes Care 35(1):S64e71.

American Diabetes Association (2010) Diagnosis and Classification of Diabetes Mellitus. Diabetes Care 33 (1).

Andrews G, Cuijpers P, Craske MG, McEvoy P, Titov N (2010) Computer therapy for the anxiety and depressive disorders is effective, acceptable and practical health care: a meta-analysis. PLoS One10:e13196.

Andrew JM, Boulton AI, Vinik JC, Arezzo VB, Eva L. Feldman R et al (2005) Diabetic Neuropathies. Diabetes Care 28(4): 956-962.

Aughsteen AA, Abu-Umair MS, Mahmoud SA (2005) Biochemical analysis of serum pancreatic amylase and lipase enzymes in patients with type 1 and type 2 diabetes mellitus. Saudi Med J 26(1):73-7.

Bader MS (2008) Diabetic foot infection. AmFam Physician 78: 71–79.

Ballinger SW, Shoffner JM, Hedaya EV, Trounce I, Polak MA, Wallace DC (1992) Maternally transmitted diabetes and deafness associated with a 10.4 kb mito-chondrial DNA deletion. Nat Genet 1:11–15.

Bamanika SA, Bamanikar AA, Arora A (2016) Study of Serum urea and Creatinine in Diabetic and non-diabetic patients in in a tertiary teaching hospital. J Med Res 2(1): 12-15.

Barroso I, Luan J, Middelberg RPS (2003) Candidate gene association study in type 2 diabetes indicates a role for genes involved in B-Cell function as well as insulin ion. PLoS Biol 1: 41-55.

Bartoov-Shifman R, Hertz R, Wang H, Wollheim CB, Bar-Tana J, Walker MD (2002) Activation of the insulin gene promoter through a direct effect of hepatocyte nuclear factor 4α. J BiolChem 277:25914–25919.

Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry. 5th edition. New York: W H Freeman; chapter 3, Protein Structure and Function.

Berke SJS, Paulson HL (2003) Protein aggregation and the ubiquitin proteasome pathway: gaining the upper hand on neuro-degeneration. Curr Opin Genet Dev 13: 253–261.

Biochemden (2018) Assay of salivary amylase activity. https://www.biochemden.com/salivary-amylase-enzyme-activity-assay/

Boj SF, Parrizas M, Maestro MA, Ferrer J (2001) A transcription factor regulatory circuit in differentiated pancreatic cells. Proc Natl Acad Sci USA 98(25):14481–14486.

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

Bonnefond A (2010) The emerging genetics of type 2 diabetes. Trends Mol Med 16(9):407-16.

Bovenschen HJ, Vissers WH (2009) Primary hemochromatosis presented by porphyria cutaneatarda: A case report. Cases J 2:7246.

Bowers LD. Kinetic serum creatinine assays (1980) The role of various factors in determining specificity. Clin Chem 26: 551-4.

Bravard A, Lefai E, Meugnier E, Pesenti S, Disse E, Vouillarmet J, Peretti N (2011) FTO Is Increased in Muscle During Type 2 Diabetes, andIts Overexpression in Myotubes Alters Insulin Signaling,Enhances Lipogenesis and ROS Production, and Induces Mitochondrial Dysfunction. Diabetes :60.

Butler AE, Dhawan S, Hoang J, Cory M, Zeng K, Fritsch H, et al (2016) β-cell deficit in obese type 2 diabetes, a minor role of β-cell dedifferentiation and degranulation. J Clin Endocrinol Metab 101:523–32.

Byrne MM, Sturis J, Menzel S, Yamagata K, Fajans SS, Dronsfield MJ et al (1996) Altered insulin secretory response to glucose in diabetic and nondiabetic subjects with mutations in the diabetes susceptibility gene MODY 3 on chromosome 20. Diabetes 45:1503–10.

Capell P (2004) Hemochromatosis in type 2 Diabetes. Clin Diabetes 22:101-2.

Carlsson S, Midthjell K, Grill V(2004) Smoking is associated with an increased risk of type 2 diabetes but a decreased risk of autoimmune diabetes in adults: an 11-year follow-up of incidence of diabetes in the Nord-Trondelag study. Diabetologia 47:1953–6.

Ceriello A (2000) Oxidative stress and glycemic regulation. Metabolism 49:27-29.

Registrar General & Census Commissioner, India (2011) Chapter 2 Size, Growth, Rate and Rural-urban distribution of population in Mizoram. Retrieved 20 June 2014.

Chen L, Magliano DJ, Zimmet PZ (2011) The worldwide epidemiology of type 2 diabetes mellitus – present and future perspectives. Nat Rev Endocrinol 8 (4): 228-36.

Chin HT (2010) Betel nut chewing and incidence of newly diagnosed type 2 diabetes mellitus in Taiwan. BMC Research Notes 3:228.

Chiu CJ, Wray LA (2011) Gender differences in functional limitations in adults living with type 2 diabetes: bio behavioral and psychosocial mediators. Ann Behav Med 41(1):71–82.

Choo-Kang A, Lynn S, Taylor G, Daly ME, Sihota SS, Wardell TM et al (2002) Defining the importance of mitochondrial gene defects in maternally inheriteddiabetes by sequencing the entire mitochondrial genome. Diabetes 51:2317–2320.

Clement K, Pueyo ME, Vaxillaire M, Rakotoambinina B, Thuillier F, Passa P et al (1996) Assessment of insulin sensitivity in glucokinase–deficient subjects. Diabetologia 39: 82–90.

Clifford J B (2017) Metformin: historical overview. Diabetologia 60:1566–1576.

Cooper RG (2006) Effect of tobacco smoking on renal function. Indian J Med Res 124, 261-268.

Cornelis M C, Qi L, Zhang C et al (2009) Joint effects of common genetic variants on the risk for type 2 diabetes in US men and women of European ancestry. Ann Intern Med 150(8) 541–550.

Coyne T, Ibiebele TI, Baade PD, Dobson A,McClintock C, Dunn S, Leonard D, Shaw J (2005) Diabetes mellitus and serum carotenoids: findings of a population based study in Queensland, Australia. Am J ClinNutr 82:685-693.

Dalgaard LT, Hansen T, Urhammer SA, Drivsholm T, Borch-Johnsen K, Pedersen O (2001) The uncoupling protein 3 -55 C -->T variant is not associated with Type II diabetes mellitus in Danish subjects. Diabetologia 44(8):1065-7.

de Courten M, Bennett PH, Tuomilehto J, Zimmet P (1997) Epidemiology of NIDDM in Non–Europids. In: Alberti KGMM, Zimmet P, DeFronzo RA, eds. International Textbook of Diabetes Mellitus. 2nd edn.Chichester: John Wiley 143–70.

de Souza BM, Brondani LA, Boucas AP, Sortica DA, Kramer CK, Canani LH et al (2013) Associations between UCP1 -3826A/G, UCP2 -866G/A, Ala55Val and Ins/Del, and UCP3 -55C/T polymorphisms and susceptibility to type 2 diabetes mellitus: case-control study and meta-analysis. PloS one 8(1):e54259.

Eliasson B (2003) Cigarette smoking and diabetes. Prog Cardiovasc Dis 45:405–13.

El Din UAAS, Salem MM, Abdulazim DO (2017) Recent Advances in Management of Diabetic Nephropathy. J Clin Exp Nephrol 2(2): 35.

Elwing JE, Gao F, Davidson NO, et al (2006) Type 2 diabetes mellitus: the impact on colorectal adenoma risk in women. Am J Gastroenterol 101(8): 1866–1871.

Facchini FS, Hollenbeck CB, Jeppesen J, Chen YD, Reaven GM (1992) Insulin resistance and cigarette smoking. Lancet 339:1128–30.

Fanglian He (2011) Bradford Protein Assay DOI: 10.21769/BioProtoc.45

Feinstein JS (1993) The relationship between socioeconomic status and health: a review of the literature. Milbank Q 71(2):279-322.

Florez JC, Hirschhorn J, AltshulerDThe (2003) basis of diabetes mellitus: implications for the genetic analysis of complex traits. Annu Rev Genomics Hum Genet 4: 257-291.

Fong DS, Aiello L, Gardner TW et al (2004) Retinopathy in diabetes. Diabetes Care 27(1): S84–87.

Francisco JP, Guillermo EU (2014) Hyperosmolar Hyperglycemic State: A Historic Review of the Clinical Presentation, Diagnosis, and Treatment. Diabetes Care 14(37): 3124–3131.

Frederick G B (1991) isn't it a sham. JAMA 198:6.

Franks PW, Pearson E, Florez JC (2013) Gene-Environment and Gene-Treatment Interactions in Type 2 Diabetes. Diabetes Care 36(5): 1413-1421.

Frayling TM, Cauchi S (2012) Stratifying type 2 diabetes cases by BMI identifies genetic risk variants in LAMA1 and enrichment for risk variants in lean compared to obese cases. PLoS Genet 8(5):e1002741.

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

Galloway JA, Chance RE (1994) Improving Insulin Therapy: Achievements and Challenges. Horm Metab Res 26(12):591-598.

Ghatak S, Muthukumaran RB, Nachimuthu S K (2013) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech 24(4), 224-231.

Gibson G (2011) Rare and common variants: twenty arguments. Nat Rev Genet 13: 135-145.

Gillespie L, Gillespie W, Robertson M, Lamb S, Cumming R, Rowe B (2006) Interventions for preventing falls in elderly people. Cochrane Database Syst Rev (3): CD000340.

Gloyn AL, Braun M, Rorsman P (2009) Type 2 diabetes susceptibility gene TCF7L2 and its role in beta-cell function. Diabetes 58(4):800–802.

Gonzalez-Salinas R, Garcia-Gutierrez MC, Garcia-Aguirre G, et al (2017) Evaluation of VEGF gene polymorphisms and proliferative diabetic retinopathy in Mexican population. Int J Ophthalmol 10(1):135–139.

Gwilt PR, Nahhas RR, Tracewell WG (1991) The effects of diabetes mellitus on pharmacokinetics and pharmacodynamics in humans. Clin Pharmacokinet 20: 477-490.

Haghvirdizadeh P, Mohamed Z, Abdullah AN, Haghvirdizadeh P, Haerian SM, Haerian SB (2015) KCNJ11: Genetic Polymorphisms and Risk of Diabetes Mellitus. J. Diabetes Res 9:908152.

Hanson RL, Millis MP, Young NJ, Kobes S, Nelson RG, Knowler WC, et al (2010) ELMO1 variants and susceptibility to diabetic nephropathy in American Indians. Mol Genet Metab 101(4):383–90.

Harris MI, Cowie CC, Stern MP, Boyko ES, ReiberGE,Bennett PH (1995) Diabetes in America. 2nd edn.Washington DC: US G.P.O (NIH publ. No. 95–1468).

Hattersley AT (2007) Prime suspect: the TCF7L2 gene and type 2 diabetes risk. J Clin Invest 117(8):2077–2079.

Henskens YM, van den Keijbus PA, Veerman EC, Van der WeijdenGA,Timmerman MF, Snoek CM (1996) Protein composition of whole and parotid saliva in

healthy and periodontitis subjects. Determination of cystatins, albumin, amylase and IgA. J Periodontal Res 31:57 65.

Homer N, Merriman B. TMAP: the T.M.A.P. [https://github.com/iontorrent/TMAP]

Hooper L, Summerbell CD, Higgins JPT, et al (2001) Dietary fat intake and prevention of cardiovascular disease: systematic review. BMJ 322(7289):757–763.

Horikawa Y, Oda N, Cox NJ, Li X, Orho-Melander M, et al. (2000) Genetic variation in the gene encoding calpain-10 is associated with type 2 diabetes mellitus. Nat Genet 26: 163–175.

Hubert K, Stephan M (2017) Environmental/lifestyle factors in the pathogenesis and prevention of type 2 diabetes. BMC Medicine 15:131.

Huyghe JR (2013) Exome array analysis identifies new loci and low-frequency variants influencing insulin processing and secretion. Nat Genet 45(2):197-2.

Ilian JGG, Yeminia V, Fernando R, Luis E FV, José FMV, Hector EFS, Bianca EGA, Nory ODR, Jorge RPG (2014) The 14 bp Del/Ins HLA-G Polymorphism Is Related withHigh Blood Pressure in Acute Coronary Syndrome and Type 2 Diabetes Mellitus. Bio Med Research 89(8):159.

Ishii D, Fujihara K, Harada M, Ishizawa M, Kodama S, Saito K, Mori Y,Honda YR, Arase Y, Sone H (2018) Association between Serum Amylase Level and Incidence of Type 2 Diabetes. Diabetes 67:1.

Jacobs N, Nicolson N, Derom C, Delespaul P, van Os J, Myin-Germeys I (2005) Electronic monitoring of salivary cortisol sapling copliance in daily life. Life Sci 76(21):2431-43.

James HB, James FG, Andrew PA,Van MS, Geoffrey BW (2004) Toward a metabolic theory of ecology. Ecology 85(7): 1771–1789.

Jiang W, Li R, Zhang Y, et al (2017) Mitochondrial DNA Mutations Associated with Type 2 Diabetes Mellitus in Chinese Uyghur Population. Sci Rep 7(1):16989.

Ji L, Malecki M, Warram JH, Yang Y, Rich SS, Krolewski AS (1997) New susceptibility locus for NIDDM is localized to human chromosome 20q. Diabetes 46:876–881.

Julia S, Esther P, Michael R (2012) The role of mitochondria in insulin resistance and type 2 diabetes mellitus. Nat Rev Endocrinol 8: 92–103.

Kai Wang, Mingyao Li, Hakon H (2010) ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res 38(16): e164.

Kahn SE (2012) Interactions between genetic background, insulin resistance and beta-cell function. Diabetes Obes Metab 14(3):46-56.

Karen L, Michal A, Jack S (2003) Evaluation of Cox's model and logistic regression for matched case-control data with time-dependent covariates:a simulation study. Statist Med 22:3781–3794.

Kaprio J, Tuomilehto J, Koskenvuo M, Romanov K,Reunanen A, Eriksson J, Stengard J, Kesaniemi YA (1992) Concordance for type 1 (insulin-dependent) and type 2 (non-insulin-dependent) diabetes mellitus in a populationbased cohort of twins in Finland. Diabetologia 35:1060-1067.

Kasuga M (2011) KCNQ1, a susceptibility gene for type 2 diabetes J Diabetes Investig 2(6): 413–414.

Kayani K, Mohammed R and Mohiaddin H (2018) Cystic Fibrosis-Related Diabetes. Front Endocrinol 9:20.

Kiukaanniemi S, Laakso M, Louheranta A, RastasM,et al (2001) Prevention of type 2 diabetes mellitus by changes in lifestyle among subjects with impaired glucose tolerance. N Engl J Med 344:1343-1350.

Kloss-Brandstätter A, Pacher D, Schönherr S, Weissensteiner H, Binna R,Specht G, Kronenberg F (2011) HaploGrep: a fast and reliable algorithm for automatic classification of mitochondrial DNA haplogroups. Hum Mutat 32(1):25-32.

Knowler WC, Barrett-Connor E, Fowler SE, Hamman RF, Lachin JM, Walker EA, Nathan DM (2002) Reduction in the incidence of type 2 diabetes with lifestyle intervention or metformin. N Engl J Med 346:393-403.

Knowler WC, Nelson RG, Saad M, Bennett PH, Pettitt DJ (1993) Determinants of diabetes mellitus in the Pima Indians. Diabetes Care 16: 216–27.

Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L,Wilson R (2012). VarScan 2: Somatic mutation andcopy number alteration discovery in cancer by exome sequencing. Genome Res 22: 568-576.

Kramer U, Herder C, Sugiri D, Strassburger K, Schikowski T, Ranft U, Rathmann W (2010) Traffic-related air pollution and incident type 2 diabetes: results from the SALIA cohort study. Environ Health Perspect 118:1273- 1279.

Krauss RM (2004) Lipids and Lipoproteins in Patients with Type 2 Diabetes. Diabetes Care 27(6): 1496-1504.

Krook A, Digby J, O'Rahilly S, Zierath JR, Wallberg-Henriksson H (1998) Uncoupling protein 3 is reduced in skeletal muscle of NIDDM patients. Diabetes 47 (9):1528-31.

Kuzuya T, Madsuda A (1997) Classification of diabetes on the basis of etiologies versus degree of insulin deficiency. Diabetes Care 20:219-220.

Lalrohlui F, Thapa S, Ghatak S, Zohmingthanga J, Kumar NS (2016) Mitochondrial complex I and V gene polymorphisms in type II diabetes mellitus among high risk Mizo-Mongoloid population, Northeast India. Gene Environ 38:5.

Lee DH, Jacobs DR Jr, Porta M (2006) Could low-level background exposure to persistent organic pollutants contribute to the social burden of type 2 diabetes. J Epidemiol Community Health 60:1006-1008.

Lee HK, Song JH, Shin CS, et al (1998) Decreased mitochondrialDNA content in peripheral blood precedes the develop-ment of non-insulin-dependent diabetes mellitus. Diabetes Res Clin Pract 42: 161–167.

Leegaard A, Rus A, Kornum JB, Prahl JB, Tomsen VO et al (2011) Diabetes,glycemic control, and risk of tuberculosis. A population-based case-control study. Diabetes Care 34: 2530–2535.

Leung CC, Lam TH, Chan WM, Yew WW, Ho KS, et al. (2008) Diabetic control and risk of tuberculosis: a cohort study. Am J Epidemiol 167: 1486–1494.

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R (2000) 1000 Genome Project Data Processing Subgroup.The Sequence Alignment/Map format and SAMtools. Bioinformatics 25(16):2078-9.

Li H, Durbin R (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics 25(14):1754-60.

Lisa A, Neda RE, Michael A (2014) Sex differences in type 2 diabetes: focus on disease course and outcomes. Diabetes Metab Syndr Obes 7 409–420.

Longnecker MP, Daniels JL (2001) Environmental contaminants as etiologic factors for diabetes. Environ Health Perspect 109(6):871-876.

López ME, Colloca ME, Páez RG, Schallmach JN, Koss MA, Chervonagura A (2003) Salivary characteristics of diabetic children. Braz Dent J 14:26 31.

Love-Gregory LD, Wasson J, Ma J et al (2004) A common polymorphism in the upstream promoter region of the hepatocyte nuclear factor-4α gene on chromosome 20q is associated with type 2 diabetes and appears to contribute to the evidence for linkage in an Ashkenazi Jewish population. Diabetes 53:1134–1140.

Lu Y, Ge Y, Shi Y, Yin J, Huang Z (2013) Two polymorphisms (rs699947, rs2010963) in the VEGFA gene and diabetic retinopathy: an updated meta-analysis. BMC Ophthalmol 13:56.

Lunter G, Goodson M (2011) Stampy: a statistical algorithm for sensitive and fast mapping of Illumina sequence reads. Genome Res 21(6):936-9.

Lynn S, Wardell T, Johnson M, Chinnery PF, Daly ME, Walker M (1998) Mitochondrial diabetes: investigation and identification of a novel mutation. Diabetes 47:1800–1802.

Lyssenko V, Jonsson A, Almgren P et al (2008) Clinical risk factors, DNAvariants, and the development of type 2 diabetes. N Engl J Med 359(21): 2220–2232.

Maassen J, t'Hart L, van Essen E, Heine R, Nijpels G, Jahangir TR et al (2004) Mitochondrial diabetes: molecular mechanisms and clinical presentation. Diabetes 53:103–109.

MacCracken J, Hoel D (1997) From ants to analogues: Puzzles and promises in diabetes management. Post grad Med 101:138–40.

Mariana M, Lijun M, Barry I (2012) Freedman. Genetic and Environmental Factors Associated With Type 2 Diabetes and Diabetic Vascular Complications. Rev Diabet Stud 9:6-22.

Marshall S, Biddle S, Gorely T, Cameron N, Murdey I (2004) Relationships between media use, body fatness and physical activity in children and youth: a meta-analysis. Int J Obes Relat Metab Disord 28(10):1238-46.

McCreight LJ, Bailey CJ, Pearson ER (2016) Metformin and the gastrointestinal tract. Diabetologia 59:426–435.

Michael C, Henry PP, Mehnaz AS, Thomas LA, Lauren G (2013). Clinical Guideline: Management of Gastroparesis. Am J Gastroenterol 108(1):18–38.

Mehrabzadeh M, Pasalar P, Karimi M et al (2016) Association between ELMO1 gene polymorphisms and diabetic nephropathy in an Iranian population. J Diabetes Metab Disord 15:43.

Meigs JB, Cupples LA, Wilson PW (2000) Parental transmission of type 2 diabetes: the Framingham Offspring Study. Diabetes 49:2201-2207.

Meirhaeghe A, Amouyel P, Helbecque N, Cottel D, Otabe S, Froguel P et al (2000) An uncoupling protein 3 gene polymorphism associated with a lower risk of developing Type II diabetes and with atherogenic lipid profile in a French cohort. Diabetologia 43(11):1424-8.

Meneilly GS, Elliott T (1999) Metabolic alterations in middle-aged and elderly obese patients with type 2 diabetes. Diabetes Care 22:112-118.

Meyre D, Bouatia-Naji N, Tounian A, Samson C, Lecoeur C et al (2005) Variants of ENPP1 are associated with childhood and adult obesity and increase the risk of glucose intolerance and type 2 diabetes. Nat Genet 37: 863–867.

Misra R, Lager J (2009) Ethnic and gender differences in psychosocial factors, glycemic control, and quality of life among adult type 2 diabetic patients. J Diabetes Complicat 23(1):54–64.

Mizuno TM (2018) Fat Mass and Obesity Associated (*FTO*) Gene and Hepatic Glucose and Lipid Metabolism. Nutrients 10(11):1600.

Mohan V, Deepa R (2006) Obesity and abdominal obesity in Asian Indians. Indian J Med Res 123(5):593-96.

Murphy R, Turnbull DM, Walker M, Hattersley AT (2008) Clinical features,diagnosis and management of maternally inherited diabetes and deafnes (MIDD) associated with the 3243A > G mitochondrial point mutation. Diabet Med 25:383–399.

Murry R (2012) Porphyrins and Bile pigment. Harper's Illustrated biochemistry. 29thed ed: McGraw Hill company 2012.

Nam HC, Kyoung MK, Sung HC,Kyong SP, Hak CJ, Sung SK, Naveed S, Soo L(2015) High Blood Pressure and Its Association With Incident Diabetes Over 10 Years in the Korean Genome and Epidemiology Study (KoGES). Diabetes Care 38:1333–1338.

Nejla S, Laila BL, Amira T, Sameh S, Mohammed J, Abrar KAA, Mohamed G, Touhami M, Wassim Y A (2018) Association of VEGFA variants with altered VEGF secretion and type 2diabetes: A case-control study. Cytokine 106:29-34.

Nicola A, Manisha C (2003) The impact of ethnicity on type 2 diabetes. J Diabetes Complicat 39–58.

Niels M, Sreekumaran N (2008) Diabetes and Protein Metabolism. Diabetes 57(1): 3-4.

Ortega-Azorín C, Sorlí JV, Asensio EM, et al (2012) Associations of the FTO rs9939609 and the MC4R rs17782313 polymorphisms with type 2 diabetes are modulated by diet, being higher when adherence to the Mediterranean diet pattern is low. Cardiovasc Diabetol 11:137.

Pablos-Mendez A, BlusteinJ, Knirsch CA (1997) The role of diabetes mellitus in the higher prevalence of tuberculosis among Hispanics. Am J Public Health 87:574–579.

Parving HH, Lehnert H, Brochner-Mortensen J, Gomis R, Andersen S, Arner P (2001) The effect of irbesartan on the development of diabetic nephropathy in patients with type 2 diabetes. N Engl J Med 345:870–878.

Pattanayak AK, Bankura B, Balmiki N, Das TK, Chowdhury S, Das M (2014) Role of peroxisome proliferator activated receptor gamma gene polymorphisms in type 2 diabetes mellitus patients of West Bengal, India. J Diabetes Investig 5(2): 188–191.

Pearson JF, Bachireddy C, Shyamprasad S, Goldfine AB, Brownstein JS (2010) Association between fine particulate matter and diabetes prevalence in the U.S. Diabetes Care 33:2196-2201.

Perry JR, Voight BF, Yengo L, Amin N, Dupuis J, Ganser M (2008) A mitochondrial DNA variant at position 16189 is associatedwith type 2 diabetes mellitus in Asians. Diabetologia 51:602–608.

Petersen KF, Befroy DS, Dufour, Dziura J, Ariyan C, Rothman DL, L. DiPietro, GW Cline, Shulman GI (2003)Mitochondrial dysfunction in the elderly: possible role in insulin resistance. Science 300: 1140–1142.

Petersen KF, Dufour S, BefroyD,et al (2004) Impaired mitochondrial activity in the insulin-resistant offspring of patients withtype 2 diabetes. N Engl J Med 350: 664–671.

Porta M (2006) Persistent organic pollutants and the burden of diabetes. Lancet 368:558-559.

Poulsen P, Kyvik KO, Vaag A, Beck-Nielsen H (1999) Heritabil-ity of type II (non-insulin-dependent) diabetes mellitus and abnormal glucose tolerance--a population-based twin study. Diabetologia 42: 139.

Pradhan AD, Manson JE, Rifai N, Buring JE, Ridker PM (2001) C-Reactive Protein, Interleukin 6, and Risk of Developing Type 2 Diabetes Mellitus. JAMA 286(3):327–334.

Prokopenko I, McCarthy MI, Lindgren CM (2008) Type 2 diabetes:new genes, new understanding. Trends Genet 24:613–621.

Qing HZ, Rui XY, Feng H, De ZY, Wei XL, Shang LP(2017) Association between the PINX1 and NAT2 polymorphisms and serum lipid levels. Onco target 8(69): 114081–114094.

Rantonen PJ, Meurman JH (2000) Correlations between total protein, lysozyme, immunoglobulins, amylase, and albumin in stimulated whole saliva during daytime. Acta Odontol Scand 58:160 5.

Rao CR, Kamath VG, Shetty A, Kamath A (2011) A cross sectional analysis of obesity among a rural population in coastal southern Karnataka, India. Australas Med J 4(1):53-57.

R Core Team (2014) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. URLhttp://www.R-project.org/

Reinke LA, Stohs SJ, Rosenberg H (1978) Altered activity of hepatic mixed-function mono-oxygenase enzymes in streptozotocin-induced diabetic rats. Xenobiotica 8 (8): 611-619.

Rhee J, Inoue Y, Yoon JC et al (2003) Regulation of hepatic fasting response by PPARg coactivator-1α (PGC-1): requirement for hepatocyte nuclear factor 4α in gluconeogenesis. Proc Natl Acad Sci 100:4012–4017.

Riobello C, Gómez J, Gil-Peña H, Tranche S, Reguero JR, de la Hera JM, et al (2016) KCNQ1 gene variants in the risk for type 2 diabetes and impaired renal function in the Spanish Renastur cohort. Mol Cell Endocrinol 15(427):86-91.

Sadikot SM, Nigam A, Das S, Bajaj S, Zargar AH., Prasannakumar KM et al (2004) Comparing the ADA 1997 and the WHO 1999 criteria: Prevalence of Diabetes in India Study. Diabetes Res Clin Pr 66: 309- 315.

Salah R, Hamed BS, Pavkoviae P, Metelko E (2002) Microalbuminuria and Diabees mellitus. Diabetologia Croatica 31-4.

Salonen JT, Nyyssonen K, Tuomainen TP, Maenpaa PH, Korpela H, Kaplan GA, Lynch J, Helmrich SP, Salonen R (1995) Increased risk of non-insulin dependent diabetes mellitus at low plasma vitamin E concentrations: a four year follow up study in men. BMJ 311:1124-1127.

Sargeant LA, Wareham NJ, Bingham S, Day NE,Luben RN, Oakes S, Welch A, Khaw KT (1996) Vitamin Cand hyperglycemia in the European Prospective Investigation into Cancer - Norfolk (EPIC-Norfolk) study. Diabetes Care 23:726–732.

Schwenk RW, Vogel H, Sch¨urmann A (2013) Genetic and epigenetic control of metabolic health. Mol. Metab 2(4): 337–347.

Shalev V, Chodick G, Heymann AD, Kokia E (2005) Gender differences in healthcare utilization and medical indicators among patients with diabetes. Public Health 119(1):45–49.5.

Sharma V, Sethi I, Sharma I, Singh G, Mahajan A, Angural A et al (2017) Replication of MACF1 gene variant rs2296172 with type 2 diabetes susceptibility in the Bania population group of Punjab, India. Int J Diabetes Dev C 126:160–3.

Sharma V, Sharma I, Sethi I, Mahajan A, Singh G, Angural A et al (2017) Replication of newly identified type 2 diabetes susceptible loci in Northwest Indian population. *Diabetes* Res Clin Pr 126:160-3.

Sharma V, Sharma I, Singh G, Sethi I, Bhanwer AJS, Singh V, Rai E, Sharma S A (2017) Replication Study of ND3 rs2853826 association with Type 2 Diabetes in North Indian population. Can J biotech 1: 257.

Shields BM, Hicks S, Shepherd MH, Colclough K, Hattersley AT, Ellard S (2010) Maturity-onset diabetes of the young (MODY):how many cases are we missing?. Diabetologia 53:2504–2508.

Shrestha S, Gyawali P, Shrestha R, et al (2008) Serum Urea and Creatinine in Diabetic and non-diabetic subjects. JNAMLS 9:11-2.

Silva HPV, Ururahy MAG, Souza KSC, Loureiro MB, Oliveira YMC, Oliveira GHM et al (2016) The HLA-G 14-BP insertion/deletion polymorphism in type 1 diabetes patients. Genes Immun 17 :13–18.

Sluik D Boeing H, Li K et al (2014) Lifestyle factors and mortality risk in individuals with diabetes mellitus: are the associations different from those in individuals without diabetes. Diabetologia 57(1): 63–72.

Solini A, Muscelli E, Stignani M, Melchiorri L, Santini E, Rossi C et al (2010) Soluble human leukocyte antigen-g expression and glucose tolerance in subjects with different degrees of adiposity. J Clin Endocrinol Metab 95:3342–6.

Stride A, Ellard S, Clark P et al (2005) Cell dysfunction, insulin sensitivity, and glycosuria precede diabetes in hepatocyte nuclear factor-1 mutation carriers. Diabetes Care 28(7):1751–1756.

Stumvoll M (2004) Control of glycaemia: from molecules to men. Diabetologia, 47: 770-781.

Sun Q, Yue P, Deiuliis JA, Lumeng CN, Kampfrath T, Mikolaj MB, Cai Y, Ostrowski MC, Lu B, Parthasarathy S et al (2009) Ambient air pollution

exaggerates adipose inflammation and insulin resistance in a mouse model of diet-induced obesity. Circulation 119:538-546.

Surajit De M, Sambanduram SS, Rajendra BM et al (2018) Metagenomic analysis and the functional profiles of traditional fermented pork fat 'sa-um' of Northeast India. AMB Express 8: 1-11.

Swapan KD (2006) Genetic Epidemiology of Adult Onset Type 2 Diabetes in Asian Indian Population: Past, Present and Future. Int J Hum Genet 6(1): 1-13.

Takeuchi M, Imano H, MurakiI et al (2018) Serum creatinine levels and risk of incident type 2 diabetes mellitus or dysglycemia in middle-aged Japanese men: a retrospective cohort study. BmJ Open diab Res Care 6:e000492.

Tang YH, Pang SM, Chan MF, Yeung GS, Yeung VT (2008) Health literacy, complication awareness, and diabetic control in patients with type 2diabetes mellitus. J Adv Nurs 62(1):74–83.6.

Taniguchi H, Honnda Y (2009) Amylases in Encyclopedia of Microbiology (Third Edition) 159-173.

Tillil H, Köbberling J (1987) Age-corrected empirical genetic risk estimates for first-degree relatives of IDDM patients. Diabetes 36: 93-99.

Trinder P (1969)Quantitative determination of glucose using GOD-PAP method. Ann Clin Biochem 6:24-27.

Valle T, Tuomilehto J, Eriksson J (1997) Epidemiology of NIDDM in Europids. In: Alberti KGMM, Zimmet P,De Fronzo RA, eds. International Textbook of Diabetes Mellitus. 2nd edn.Chichester: John Wiley: pp125–42.

van den Ouweland JMW, Lemkes HHPJ, Ruitenbeek W, Sandkujli LA, deVijlder MF, Struyvenberg PAA, van de Kamp JJ, Maassen JA (1992) Mutation inmitochondrial tRNAL e u (U U R)gene in a large pedigree with maternally transmittedtype 2 diabetes mellitus and deafness. Nat Genet1:368–371.

Vigersky RA (2011) An overview of management issues in adult patients with type 2 diabetes mellitus. J Diabetes Sci Technol 5(2): 245–250.

Vijay V, Snehalatha C, Shina K (1993) Familial aggregation of diabetic kidney disease in type 2 diabetes in south India. Diabetes Res ClinPract 43:167–71.

Walker M, Turnbull DM (1997) Mitochondrial related diabetes:a clinical perspective. Diabet Med 14**:** 1007–09.

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

Watkins JB 3rd, Sanders RA, Beck LV (1988) The effect of long-term streptozotocin-induced diabetes on the hepatotoxicity of bromobenzene and carbon tetrachloride and hepatic biotransformation in rats. Toxicol. Appl Pharmacol 93: 329-338.

WD DWaP (1997) PS power and sample size program available for free on the Internet. Controlled Clin Trials 18:18:274.

Weedon MN, Owen KR, Shields B et al (2004) Common variants of the hepatocyte nuclear factor-4α P2 promoter are associated with type 2 diabetes in the U.K. population. Diabetes 53:3002–3006.

Wei L, Xiao Y, Li L, Xiong X, Han Y, Zhu X, Sun L (2018) The Susceptibility Genes in Diabetic Nephropathy. Kidney Dis 4:226-237.

Weill J, Vanderbecken S, Froguel P (2004) Understanding the rising incidence of type 2 diabetes inadolescence. Arch Dis Child 89: 502-504.

Wessel J, Brody JA, Manning A, Topmed Diabetes Working Group (2018) Whole Genome Sequence Association Analysis and Diabetes-centric Functional Annotation of Type 2 Diabetes. Diabetes 67 (1): 24-OR; DOI: 10.2337/db18-24-O.

Wieczór R, Gadomska G, Ruszkowska-Ciastek B et al (2015) Impact of type 2 diabetes on the plasma levels of vascular endothelial growth factor and its soluble receptors type 1 and type 2 in patients with peripheral arterial disease. J Zhejiang UnivSci B 16(11):948–956.

Winckler W, Weedon MN, Graham RR, McCarroll SA, Purcell S, Almgren P et al (2007) Evaluation of common variants in the six known maturity-onset diabetes of the young (MODY) genes for association with type 2 diabetes. Diabetes 56(3):685-93.

Whittaker RG, Schaefer AM, McFarland R, Taylor RW, Walker M, Turnbull DM (2007) Prevalence and progression of diabetes in mitochondrial disease. Diabetologia 50, 2085–2089.

WHO(1999) Definition, Diagnosis and Classification of Diabetes Mellitus and its Complications, Report of a WHO Consultation. Geneva, Switzerland:WHO.

WHO (2000) Obesity: Preventing and Managing the Global Epidemic,Report of a WHO Consultation.Geneva,Switzerland:WHO.

Xie X, Liu Q, Wu J, Wakui M (2009) Impact of cigarette smoking in type 2 diabetes development. Acta Pharmacol Sin 30(6): 784–787.

Zatalia SR, Sanusi H (2013) The role of antioxidants in the pathophysiology, complications, and management of diabetes mellitus. Acta Med Indones 45(2): 141–147.

Zee RY, Ridker PM, Chasman D (2011) Genetic variants of 11 telomere-pathway gene loci and the risk of incident type 2 diabetes mellitus: the Women's Genome Health Study. Atherosclerosis 218(1):144–146.

Zhang H, Zhang C, Tang H, et al (2018) CD2-Associated Protein Contributes to Hepatitis C, Virus Propagation and Steatosis by Disrupting Insulin Signaling. Hepatology 68(5):1710–1725.

Zimmet PZ (1992) Kelly West Lecture 1991: challenges in diabetes epidemiology: from West to the rest. Diabetes Care 15: 232–52.

Zouali H, Hani EH, Philippi A et al (1997) A susceptibility locus for early-onset non-insulin dependent (type 2) diabetes mellitus maps to chromosome 20q, proximal to the phosphoenol pyruvate carboxykinase gene. Hum Mol Genet 6:1401–1408.

Appendices

Appendix I: List of publications in peer-reviewed journals

- **Lalrohlui F**, Thapa S, Ghatak S, Zohmingthanga J, Senthil-Kumar N. (2016). Mitochondrial complex I and V gene polymorphisms in type II diabetes mellitus among high risk Mizo-Mongoloid population, Northeast India. *Genes and Environment* 38(1):1–10
- Thapa S, **Lalrohlui F**, Ghatak S, Zohmingthanga J, Lallawmzuali D, Pautu JL, Senthil Kumar N. (2015). Mitochondrial complex I and V gene polymorphisms associated with breast cancer in mizo-mongloid population. *Breast Cancer* 23(4):607–616
- Ghatak S, Yadav RP, **Lalrohlui F**, Chakraborty P, Ghosh S, Ghosh S, Pautu JL, Zohmingthanga J, Senthil Kumar N. (2016). Xenobiotic Pathway Gene Polymorphisms Associated with Gastric Cancer in High Risk Mizo-Mongoloid Population, Northeast India. *Helicobacter* 1523-5378.
- Ghatak S, Lalnunhlimi S, **Lalrohlui F**, Yadav RY, Chakraborty P, Pautu JL, Zohmingthanga J, Roy NK, Kunnumakkara AB, Lalhruaitluanga H, Senthil-Kumar N. (2018). Novel AKT1 mutations associated with cell cycle abnormalities in Gastric carcinoma. *Personalized Medicine* 15(2): 79–86.
- Ravi Prakash Yadav, Souvik Ghatak, Payel Chakraborty, **Freda Lalrohlui**, Ravi Kannan, Rajeev Kumar, Jeremy L Pautu, John Zomingthanga, Saia Chenkual, Rajendra Muthukumaran, Nachimuthu Senthil Kumar (2018). Lifestyle chemical carcinogens associated with mutations in cell cycle

regulatory genes increases the susceptibility to gastric cancer risk . *Environmental Science and Pollution Research* 25: 31691.

- Mary V Tonsing, Souvik Ghatak, **Freda Lalrohlui**, Nachimuthu Senthil Kumar, John Zohmingthanga. Five Years Record on Cancer Incidence from a Diagnostic Centre in Mizoram, Northeast India. *Cancer Health Disparities* 2:e1-e15. doi:10.9777/chd.2018.10014.

Manuscript Accepted

- **Freda Lalrohlui,**Varun Sharma, Indu Sharma, Sarmeela Sharma ,Guneet Kour , Shruti Sharma, Surbhi Sharma, Yuman, Tasmeen Javed Parihar, Hemender Singh, John Zohmingthanga, Ekta Rai, Vinod Singh, Nachimuthu Senthil Kumar, Swarkar Sharma. First replication study from Northeast India indicates association of variant rs2296172 of MACF1 gene with type 2 diabetes in Mizo population. **International Journal of Diabetes in Developing Countries**
- **Freda Lalrohlui,** Varun Sharma Indu Sharma, Ekta Rai, John Zohmingthanga, Vanlal hruaii,Swarkar Sharma2 and Nachimuthu Senthil Kumar. Genotyping of T2D Susceptible Genes in a High Risk North-East Indian Population. **Obesity Medicine**

- **Freda Lalrohlui,** John Zohmingthanga, Vanlal hruaii, and Nachimuthu Senthil Kumar. Genomic profiling of Mitochondrial DNA reveals novel complex gene mutations in familial type 2 diabetes mellitus individuals from Mizo ethnic population, Northeast India **.Mitochondrion**

Manuscript Under Review

- **Freda Lalrohlui**, Souvik Ghatak, John Zohmingthanga,Vanlal hruaii ,Nachimuthu Senthil Kumar. Fermented pork fat (Sa-um) and Tobacco infused water (tuibur) are the major risk factors for type 2 diabetes among Mizo population, Northeast-India. **Journal of Health, Population and Nutrition**
- **Freda Lalrohlui**, Subhajit Mukherjee, John Zohmingthanga,Nachimuthu Senthil Kumar .Saliva as potential diagnostic marker of total protein concentration for type 2 diabetes mellitus patients –A Pilot Study. **Journal of Basic and Clinical Physiology and Pharmacology**

- **Freda Lalrohlui**;Sarmeela Sharma;Varun Sharma; Indu Sharma; Shruti Sharma; Tasmeen Javed Parihar; John Zohmingthanga; Vinod Singh; Ekta Rai; Nachimuthu Senthil Kumar; Swarkar Sharma. UCP3 variant rs1800849 did not show association with T2D in Mizo population of Northeast India. **International Journal of Diabetes in Developing Countries**
- Freda Lalrohlui, John Zohmingthanga, Vanlal hruaii, Nachimuthu Senthil Kumar. Whole exome sequencing identifies the association of novel gene variants related with type 2 diabetes in Mizo population, Northeast India. **Gene.**

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

- Participated as a **Resource Person** in the **"A Primer in Bioinformatics and Biotechnology to School Students"** from June 08, 2013 at DBT-State Biotech HUB, Department of Biotechnology, Mizoram University.
- **Poster presentation** in the **" Advances in Cancer Genomics"** from May 30 – 31, 2014 Organized jointly by Mizoram State Cancer Institute and Department of Biotechnology, Mizoram University, Aizawl, Mizoram.
- Participated as a **Resource Person** in the "**Hands on training and DNA barcoding and phylogenetics**" from March 20-25, 2017 at Advanced State Biotech HUB, Department of Biotechnology, Mizoram University.
- Participated and **presented a paper** in the "**National conference on Recent Advances in Biotechnology**" from 9-10 November, 2017 at Mizoram University.
- Participated as a **Resource Person** in the "**Hands on training on Molecular Research Technique**" from 31 May to 6 June, 2018 at Govt. Zirtiri Residential Science College, Aizawl.
- Participated and **presented a poster** in the **"Annual IMA Pre-Conference Continuing Medical Education**" on 13 December ,2018 organized by Indian Medical Association, Mizoram State Branch at Aijal Club, Aizawl.
- **Poster presentation** at "**INSPIRE Fellowship Review Meet**" from May 30 – 1 June, 2019 Organized by Department of Science and Technology, New Delhi, India at IASST Garchuk, Guwahati.

Award

- Best Oral presenter in the "**National conference on Recent Advances in Biotechnology**" from 9-10 November, 2017 at Mizoram University.

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

- Training course of "**Bioinformatics-Protein and their Structure Prediction**" Sponsored by Department of Biotechnology, Ministry of Science and Technology, Government of India, Mizoram University, 23-24th November, 2011.
- Participated as a participant in the "**Structure Determination of Macromolecules**" from March 26-28, 2013 at DBT-BIF Centre, Department of Biotechnology, Mizoram University.
- Participated as a participant in the "**Workshop on Capacity Building in Effective Management of Intellectual Property Right (IPRs**)" from August 27 - 28, 2014 at Mizoram University, Aizawl, Mizoram
- Participated as a participant in the "**Cancer Mutations - Detection and Analysis**" from April 17 – 18, 2015 at DBT-State Biotech HUB, Department of Biotechnology, Mizoram University.
- Participated as a participant in the "**Browsing genome with Ensemble and UCSC genome browser**" from 30th May, 2015 at DBT-State Biotech HUB, Department of Biotechnology, Mizoram University.
- Participated as a participant in the "**Gene ontology and cytoscope**" from July 13-14 2015 at Department of Biotechnology, Mizoram University.
- Participated as a participant in the "**Sensitization Workshop on Research Methodology**" from 15-16 February, 2016 organized by Indian Council of Medical Research at Civil Hospital, Aizawl, Mizoram.

- Participated as a participant in the "**Science Communication workshop**" held on 6 June, 2017 at Mizoram University, Aizawl.
- Participated as a participant in the "**Research Training Workshop for North East Region Scientists on Understanding Human Disease and Improving Human Health Using Genomics-Driven Approaches**" organized by the National Institute of Biomedical Genomics (NIBMG), Kalyani from November 19-24, 2017 at Mizoram University, Aizawl
- Participated as a participant in the "**The Concept and Application of Genomics in Clinical Medicine**" on 8 August, 2018 organized by Civil Hospital, Aizawl and Mizoram University at Civil Hospital, Aizawl, Mizoram.
- Participated as a participant in the "**12th Annual Convention of Association of Biotechnology and Pharmacy (ABAP) & International Conference on Biodiversity, Environment and Human Health: Innovations and Emerging Trends (BEHIET**)" from 12-14 November, 2018 organized by the School of Life Sciences, Mizoram University.

IEC, CIVIL HOSPITAL, AIZAWL.

COMMUNICATION TO THE PRINCIPAL INVESTIGATOR/GUIDE BY THE MEMBER SECRETARY, INSTITUTIONAL ETHICS COMMITTEE

No.B.12018/1/13-CH(A)/IEC/36 **Dated: 23rd December 2015**

To,

Dr. N. Senthil Kumar,
Professor & Head
Department of Bio-technology,
Mizoram University

Subject: Ethics Committee Approval for the referenced projects.

Reference: **"Genetic Perspectives of Type 2 Diabetes Susceptibility in Mizo Population, Mizoram, North East India."**

Dear Dr. N. Senthil Kumar,

With reference to submission of document for review and approval to conduct the above mentioned study.

The Ethics Committee has reviewed and approved the study documents as mentioned below:

1. Curriculum Vitae of non Civil Hospital, Aizawl Investigators
2. Brief description of proposal/summary
3. Copy of the Protocol/Project and questionnaire (if any)
4. Copy of Patient information sheet & Consent form in local language
5. Copy of Clinical trial agreement
6. Copy of PI undertaking

The following members of the Ethics Committee were present at the meeting held on date **22nd December 2015** at time **1:00pm** at Library Room, Civil Hospital, Dawrpui, Aizawl, Mizoram - 796001.

The quorum met as per ICG-GCP and schedule Y guidelines as mentioned below:

S. No	Name of the member	Qualification	Designation in the Ethics Committee	Gender
1.	Dr. L.Ringluaia DRCOG	Retired Director, Synod Hospital	Chairman	Male
2.	Dr. C.Lalchhandama MD	Senior Pathologist	Member Secretary	Male
3.	Dr. H.C. Laldina, MD	Head of Dept, ENT	Member	Male
4.	Pu Rosanglura Ralte,	Advocate	Legal expert	Male
5.	Dr. T. Lalzawmliana, MD	Consultant, Dept of Biochemistry	Basic Medical Science	Male

Please note that this Ethics Committee is constituted as per schedule Y, ICH-GCP, applicable local laws and regulatory requirement.

We approve the project to be conducted in its present form. Ethics Committee expects to be informed about:

1. Any SAE occurring in the course of the study
2. A copy of final individual center report

We hereby confirm that neither you nor your study members have participated in the voting/decision making procedure of the Ethics Committee.

Yours sincerely,

Sd/-

(DR. L.RINGLUAIA)
Chairman
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram – 796001

Chairman
Institutional Ethical Committee
Civil Hospital, Aizawl

23/12/15

(DR. C.LALCHHANDAMA)
Member Secretary
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram - 796001

Secretary
Institutional Ethical Committee
Civil Hospital, Aizawl

ANNEXURE - VII

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

No.B.12018/1/13-CH(A)/IEC/36 **Dated: 23rd December 2015**

To,

Dr. N. Senthil Kumar,
Professor & Head
Department of Bio-technology,
Mizoram University

Subject: Ethics Committee Approval for the referenced projects.

Reference: **"Genetic Perspectives of Type 2 Diabetes Susceptibility in Mizo Population, Mizoram, North East India."**

Dear Dr. N. Senthil Kumar,

With reference to submission of document for review and approval to conduct the above mentioned study. The Ethics Committee has reviewed and approved the study documents as mentioned below:

1. Curriculum Vitae of non Civil Hospital, Aizawl Investigators
2. Brief description of proposal/summary
3. Copy of the Protocol/Project and questionnaire (if any)
4. Copy of Patient information sheet & Consent form in local language
5. Copy of Clinical trial agreement
6. Copy of PI undertaking

The following members of the Ethics Committee were present at the meeting held on date **22nd December 2015** at time **1:00pm** at Library Room, Civil Hospital, Dawrpui, Aizawl, Mizoram - 796001.

The quorum met as per ICG-GCP and schedule Y guidelines as mentioned below:

S. No	Name of the member	Qualification	Designation in the Ethics Committee	Gender
1.	Dr. L.Ringluaia DRCOG	Retired Director, Synod Hospital	Chairman	Male
2.	Dr. C.Lalchhandama MD	Senior Pathologist	Member Secretary	Male
3.	Dr. H.C. Laldina, MD	Head of Dept, ENT	Member	Male
4.	Pu Rosanglura Ralte,	Advocate	Legal expert	Male
5.	Dr. T. Lalzawmliana, MD	Consultant, Dept of Biochemistry	Basic Medical Science	Male

Please note that this Ethics Committee is constituted as per schedule Y, ICH-GCP, applicable local laws and regulatory requirement.

We approve the project to be conducted in its present form. Ethics Committee expects to be informed about:

1. Any SAE occurring in the course of the study
2. A copy of final individual center report

We hereby confirm that neither you nor your study members have participated in the voting/decision making procedure of the Ethics Committee.

Yours sincerely,

Sd/-

(DR. L.RINGLUAIA)
Chairman
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram – 796001

Chairman
Institutional Ethical Committee
Civil Hospital, Aizawl

23/12/15

(DR. C.LALCHHANDAMA)
Member Secretary
Institutional Ethics Committee
Civil Hospital, Dawrpui, Aizawl
Mizoram - 796001

Secretary
Institutional Ethical Committee
Civil Hospital, Aizawl

PATIENT PERFORMA FOR STUDY FOR TYPE 2 DIABETES MELLITUS

Department of biotechnology
Mizoram University
Tanhril,Mizoram

Patient Data Sheet

1. Serial No.**:** Date:

2. **Profile**

Name of the Patient(HMING):

Caste/ Tribe -

Age(Kum)**:** Sex: Male / Female

Address: Ph. No:

Background:

Job/profession:

3 Duration of Hyperglycemia and year in which diagnosed(Thisen thlum/Zun thlum hi kum engzat nge I vei tawh?Egtik kuma hmuh chhuah nge a nih?**):**

_____

4.Smoking/ Tobacco in any other form. (Meizial emaw vaihlo lampang telna chi hrim hrim I ti m?)(YES/NO)

If Yes, which type of tobacco product?(I tih chuan eng lampang vaihlo telna nge?) : -

5. Hypertensive(Thisen sang I nei em ?):- Yes / No

6. Food habits (Ei leh in chungchang):

Do you consume non-veg? (Sa I ei em?) YES ( ) No ( )

(a)Item consumed (Thil ei): Meat/Sa ( ) Fish/sangha ( ) Egg /artui (boiled/fried/cooked)

How frequent ?(Engtianga zing nge?) _____

(b)How do you normally consume (engtiangin nge I ei thin ?):
(Chhum/boiled)/ (Kan/fried)/ cooked foods

(c)Which type of food do you prefer?(Heng zing atang hian eng hi nge I ei duh ber?)

red meat(sa hring) _____
smoked meat(sa rep) _____
Processed food(sa/chaw sawngbawl sa) _____

Salt intake(Chi eidan) : Heavy ( ) average ( ) little ( )

Fat intake(Thau/mawm ei tam lam): heavy ( ) average ( ) little ( )
_____

Fruits consumption(thei ei tam dan) – Regular/ Not regular
(zing/zinglo)

How often do you do any exercise?(exercise I la zing em?)

( ) rarely/never ( ) 2-3 times a week
( ) less than once a week ( ) 4-6 times a week
( ) once a week ( ) every day

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

7.Do you undergo pancreatic surgery ?(YES/NO).If YES,When ?_____

FAMILY HISTORY

8. Persons having Diabetes in the family (Chhungkhata zunthlum nei?)

Father ( ) Mother ( ) sister ( ) brother ( ) 2° ( ) others_____

Family persons having stroke or cardiovascular disease_____
(Chhungkuaah lung lam natna emaw stroke emaw nei tawh?)

TREATMENT HISTORY

What type of Medication do you take ? (Eng ang damdawi nge I lak?)

(a)Oral_____(ei chi)

(b)Injection _____ (inchiu chi)

(c) Others _____(a dangte)

Duration of treatment_____ (treatment lak rei zawng)

- **Examination** :-

BMI (Weight/Height²):- **Kg/Cm²** **BP: -** **mm/Hg**

Ankle: Brachial pressure index:-

WHR (Waist Hip Ratio):- **HbA1c:-**

Dysepidaemic: - Y / N ECG:-

Evidence of Retinopathy: - Evidence of Neuropathy:-

Evidence of -IHD----/ peripheral arterial disease / cerebrovascular disease:-

Fasting sugar: - **PP sugar:-**

Triglyceride: - LDL:-

HDL: - VLDL:-

Cholesterol: - Creatinine:-

Donor consent form (Sample lak phalna)

I have gone through the information regarding the research study, I am willing to donate my blood sample to be utilized for medical research. I am aware that my acceptance or rejection to allow my sample being utilized for the above mentioned research study is not going to effect my treatment or prognosis. I also understand that I can withdraw my sample involvement in the research study at any point of time. The blood will be drawn by an experienced phlebotomist.*This study will direct us to the indepth of the causes, prevalence and factors contributing to Type II Diabetes Mellitus*.

The samples will be made available for researchers at MZU (department of Biotechnology, Mizoram University), and other collaborating universities, institutes, hospitals/ clinics. The result will be maintained in confidentiality and will not be disclosed to any third party who are not involved in this scientific research. The information about the results/findings of the researcher may not be directly given to the patient.

(*He research tihna tur hi ka zircian tawh hnu ah ka thisen sample hi medical research atan lak ka phal ngei a ni.He research tihna tana ka sample hman ka phal hian ka in enkawlna zel leh ka that chhohna zelah nghawng a nei dawnlo tih ka hria e.Ka sample hi hman ka phallo anih chuan eng hunah pawh ka in hnuk dawk thei e.Thisen hi thisen lak lama thiamna neiin a la ang.He zir chianna hian Type II Diabetes Mellitus chungchang thuk zawk te,ciang zawk te leh eng vanga hluar ta viau nge ani tih a hriat theih dawn ani.*

Sample hi MZU a zirmi te leh an thurualpui University,hospital,clinic ho hman theihin a awm anga.He thil zircianna atanga hmuh chhuah hi he thil a tel velo hnenah hmuh emaw hriattir a ni lo ang.Result hmuhchhuah hi damlo te hnenah direct in hrlh emw hriattir an ni lo ang)

Signature of donor Date/Time

Signature of Investigator Date/Time

Researcher's contact information
(heng a hnuaia number leh email ah hian kan biak pawh theih reng e)

Freda Lalrohlui: 08794022599 email:fredaxsmart@gmail.com

Prof N.Senthil Kumar: 09436352574 email:nskmzu@gmail.com

Dr John Zohmingthanga : 09436140552 email:johnzo05t@yahoo.co.in

Information obtained after research study

Mutation detected: **Gene / Exon:**

Type of Mutation:

Codon Position / Amino acid change:

Familial history / Pedigree:

Lalrohlui et al. Genes and Environment (2016) 38:5
DOI 10.1186/s41021-016-0034-z

Genes and Environment

SHORT REPORT Open Access

Mitochondrial complex I and V gene polymorphisms in type II diabetes mellitus among high risk Mizo-Mongoloid population, Northeast India

Freda Lalrohlui[1], Sunaina Thapa[1], Souvik Ghatak[1], John Zohmingthanga[2] and Nachimuthu Senthil Kumar[1*]

Abstract

Introduction: The study was carried out to identify the polymorphisms in mitochondrial genes (ATPase and ND1) in type 2 Diabetes Mellitus (T2DM) from Mizo population and to correlate the involvement of demographic factors.

Findings: In the present study, 58 patients and 50 healthy volunteers were considered. The mutations observed were mostly base substitutions and were similar as reported for other populations. Three mutations are unreported and were found to be novel polymorphisms for diabetic disease. One heteroplasmic variation (MT3970 C > T) was found in 36.36 % of samples. Subjects with excessive smoked meat consumption and customary habit of smoking (ORs: 4.92; 95 % CI: 0.96–25.21) were found to be more prone to T2DM. Mitochondrial genes sequence analysis revealed the genetic variability between the healthy and diabetic samples.

Conclusion: Mitochondrial ATPase and ND1 gene polymorphisms may be involved in triggering the risk for T2DM.

Keywords: Type 2 diabetes mellitus, Demography, Polymorphism, Mitochondrial genes, Genetic variability

Introduction

Mitochondrial dysfunctions are involved in ageing and age-related diseases such as Diabetes [1]. Complex I and V is one of several enzyme complexes necessary for oxidative phosphorylation [2]. Patients with large mtDNA mutations like deletion, deletion-duplication or in association with mtDNA point mutations generally in tRNA genes (tRNA (LEU(UUR)) has been reported with Diabetes Mellitus [3]. Na+,K + −ATPase is an ubiquitous membrane enzyme that allows the extrusion of three sodium ions from the cell and two potassium ions from the extracellular fluid. Abnormal accumulation of ROS activates UCP2, which in turn results in proton leak across the mitochondrial inner membrane leading to reduced b-cell ATP synthesis and content. This is a critical parameter in regulating glucosestimulated insulin secretion and release which ultimately increases circulating blood glucose level [4]. ND1 gene provides directives for making a protein called NADH dehydrogenase I. The actions of mitochondrial content are often reduced in patients with T2DM, or insulin resistance [5, 6]. Type II diabetes (T2D) is considered as the heterogeneous disease with altered insulin production by the pancreatic beta cell. The study of the relationship of ATPase and ND1 gene to type 2 diabetes has revealed the influence of the mitochondria on nuclear-encoded glucose transporters and the influence of nuclear encoded uncoupling proteins on the mitochondria [7]. There is evidence of a more global effect of mitochondrial dysfunction at the glucose transporter level and it will be interesting to study the variations in the genes involved among the different populations.

Earlier studies showed that mtDNA ND1 gene mutations at nt3310 (C > T), nt3667 (T > G) might contribute to the pathogenesis of DM with other genetic factors and environment factors [8, 9]. Howarth and Worsley [9] studied the dietary habits of elderly diabetics and have shown that

* Correspondence: nskmzu@gmail.com
[1]Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India
Full list of author information is available at the end of the article

faulty diet regimes can make the best of medicine ineffective.

Mizoram is one of the northeastern states of India, bordered by Bangladesh in the west and Myanmar on the east and south. Mizo people belong to the Mongoloid race and are ethnically and culturally most diverse tribe in the world [10]. Mizoram lies between 21°58′ & 24°35′ N latitude and 92°15′ & 93°29′E longitude and spread over 21,081 sq. kms area. Traditional Mizo food mostly comprises of boiled, stewed, smoked, steamed, or fermented form. Mizo food also comprises of certain leafy vegetables, fresh as well as preserved through smoking, such as mustard leaves (antam), pumpkin leaves (maian), beans leaves (behlawi), varieties of bamboo shoot (mautuai, rawtuai), fermented soya beans (bekang), fermented lard (sa-um) and dried fish chutney with green chilly. A peculiar habit of consumption of "tuibur" (tobacco smoke–infused aqueous solution) has been observed in Mizoram [11]. The present study was carried out to understand the influence of demographic factors on type 2 diabetics and associated mitochondrial polymorphisms in the Mizopopulation, North-east India.

Materials and methods

Sample collection and DNA extraction

A total of 58 patients with or without a family history of type 2 Diabetes Mellitus (median age 48 years, range 24–77) from Civil Hospital, Aizawl, Mizoram, India and 50 healthy volunteers (median age 48 years, range 35–63) were randomly recruited for this study from Mizo population. Senior diabetologist confirmed the diagnosis of Type 2 diabetes mellitus. The peripheral blood samples of these affected patients were kept in EDTA rinsed microcentrifuge tubes and stored in −20 °C freezer. Detailed information on demographic factors such as physical activity, dietary habits, previous disease history, alcohol and tobacco use and family history of diabetes were recorded during an in-person interview using a structured questionnaire (Additional file 1: Table S1). The ethical committee of all institutes approved the study protocol involved in the study. All volunteers were fully informed about the study and participated with their full consent. DNA extraction and quantification from the blood samples were performed according to Ghatak et al. [12].

PCR amplification of ATPase and NDI gene

The mtDNA ATPase region (1046 kb) was amplified by PCR using primers KatPase-F (5′-CTAGAGCCCACTG TAAAGCTAAC-3′) and KatPase-R(5′-GAGCGTTATG GAGTGGAA GT-3′). Polymerase chain reaction (PCR) for ATPase region was carried out in 25 μl total reaction volume, each containing 100 ng of template DNA, 0.25 pM of each primer, 2.5 μl of 10X PCR buffer, 1.5 mM $MgCl_2$, 200 mM dNTPs, and 1.5 U of Dream Taq green DNA polymerase (Fermentas, Germany). Polymerase chain reaction volume was heated initially to 95 °C for 5 mins. followed by 30 cycles each consisting of 1 min. denaturation at 95 °C, 40 s annealing at 56 °C, 1 min. extension at 72 °C. The reaction ended with a final extension step by incubating at 72 °C for 5 min. The entire mitochondrial ND1 gene spanning across nucleotide positions 3306 to 4261 was amplified. PCR amplification was performed using forward primer (5′-GAGCCCGG TAATCGCATAA-3′) and reverse primer (5′-GATAGG TGGCACGGAGAAT-3′). Polymerase chain reaction (PCR) for ND1 gene region was carried out in 25 μl total reaction volumes, each containing 100 ng of template DNA, 0.2 pM of each primer, 2.5 μl of 10X PCR buffer, 2 mM $MgCl_2$, 200 mM dNTPs and 1 U of Dream Taq green DNA polymerase (Fermentas, Germany). The mixture was subjected to initial denaturation at 95 °C for 5 min., followed by 30 cycles of denaturation at 95 °C for 60 s, annealing of primers at 58 °C for 40 s, extension at 72 °C for 70 s. and a final extension cycle at 72 °C for 5 min. The PCR products were subjected to electrophoresis in a 1.2 % Agarose gel in 1X TBE buffer, stained with Ethidium Bromide, and images were obtained in GBOX gel documentation systems (UK) and sequenced.

RFLP of PCR amplified product

The PCR amplified products were subjected to digestion using *AciI* (8–10 h at 37 °C), *Hae III* (3 h at 37 °C), *TaqI* (10–12 h at 56 °C) and *RsaI* (6–10 h at 37 °C in a total volume of 10 μl containing 3 μl of DNA, 0.4 μl of enzyme, 1 μl of buffer and 5.6 μl of water. The digested products were subjected to electrophoresis in 8 % PAGE (Polyacrylamide gel electrophoresis) gel at 40 V for 30 min and changed to 60 V, post staining was done using 1 μl of ethidium bromide and images were obtained in GBOX gel documentation systems.

Sequence analysis

Based on the above digestion experiment, the polymorphic samples were selected and sequenced from both directions to ensure reading accuracy. Sequences and chromatograms obtained were examined using FINCH TV 1.4 software version (Geospiza. Inc., USA) and DNA Baser software version 4.16 and aligned by BLAST (http://www.ncbi.nlm.nih.gov/blast). All sequences were comparedwith the latest version of Revised Cambridge Reference Sequence (rCRS) and subsequently analyzed for the variation in sequences using Mito Tool Programming. The results of the DNA sequence analysis were compared with the published Cambridge Sequence using Mutation Surveyor version 1.4 DNA mutation analysis software (Softgenetics, State College, PA). Sequence differences between diabetic and healthy blood samples were recorded as mtDNA polymorphisms. Each polymorphism was

verified against the Mitomap database (http://www.mitomap.org/) and further classified as novel or reported, depending on whether or not it is recorded in the database. The effect of amino acid substitutions based on the single nucleotide positions were predicted using Polyphen 2 software.

The number of base substitutions per site between sequences and averaging over all sequence pairs within each group were analyzed. Analysis was conducted using the Tamura 3-parameter model. The rate variation among sites were modeled with a gamma distribution (shape parameter = 1). The analysis involved 15 nucleotide sequences. Analyses were conducted in MEGA6. The variable substitution site was calculated by DAMBE: Software Package [13].

Statistical analysis

Hardy-Weinberg equilibrium by a chi-square ($\chi 2$) test with one degree of freedom (df) was performed between case and control subjects. Fisher's exact test was also used for comparing the demographic and habits between patients and controls. The polymorphism and demographic factor in each group were estimated for their association with diabetes using odds ratios (ORs) and 95 % confidence intervals (CIs) in the Logistic Regression (LR) Model adjusted with multivariable analysis. Each polymorphism was checked by the presence and absence of the SNPs. Additionally, logistic regression analyses were conducted to compute the influence of both genetic and environmental factors. For all tests, a two-sided *P*-value <0.05 was considered statistically significant. All statistical analyses were performed using SPSS 20.0 program (SPSS Ibérica, Madrid, Spain) and SYSTAT 13.0. (Systat Software Inc., USA).

Results

In the present study, blood samples of 58 Diabetic patients and 50 healthy individuals were analyzed. The prevalence of Type 2 diabetes was higher in patients who consumed smoked meat and excess fat (Odd Ratio, OR: 4.76, 95 % Confidence Interval, 95 % CI: 1.03–13.73). This was observed especially among men (OR: 1.14, 95 % CI: 0.25–1.63). Smoking, consuming betel-nut with paan and alcohol were major risk factors for type 2 diabetes. In Mizo population, Type 2 diabetes is not significantly associated with familial history. There are significant differences in the age of onset ($p < 0.001$) of diabetes and there was no significant differences in gender or concentration of plasma glucose level (Additional file 1: Table S1). As is typical for Hardy-Weinberg equilibrium, the degree of significance was quite coarse due to the less number of sample size using both chi-square ($\chi 2$) and fisher's exact tests.

Total of 6 ATPase (ATP6 and ATP 8) and ND1 sequence variations at six distinct nucleotide positions were found in 21.8 and 54.54 % samples, respectively. One heteroplasmic variation (MT3970 C > T) was found in 36.36 % of samples (EBI Accession No. LN558438 - LN558467and NDI regions - LN558501 - LN558513). In the ATPase and ND1 genes, 6 non-synonymous and 6 synonymous substitutions were found and out of these, 3 non-synonymous and 2 synonymous were not previously reported in the literature or the public mtDNA mutation databases (mtDBase:http://www.genpat.uu.se/mtDB/index.html; MITOMAP:http://mitomap.org/ MITOMAP) related with diabetes (Table 1).

Genetic divergence was estimated between the diabetic and healthy samples by calculating the number of base substitutions. Divergence distance between the diabetic and healthy samples were found to be 0.002 for ATP6, nil for ATP 8 and 0.002 for ND1 genes. The diabetic and healthy samples clustered in two separate clades for ATP6 and ND1 gene sequences (data not shown). Majority of the nucleotide diversity were present in the start region for ATP6 and both start and end regions for NDI gene. The transition for all the three genes is higher than transversion. The transition- transversion bias for the ATP6 is high for diabetes and healthy control samples (Table 2). The haplotype frequency P, Watterson's Θ, which compares two estimators of the population parameter is higher in ATP8 gene than NDI and ATP8. Based on the Tajima test statistics, ATP6 gene is a good marker for genetic variation analysis between the healthy and diabetic samples (Table 2).

Discussion

The main objective of the study was to find out the role of demographic factors in the onset of diabetes and to detect their associated mutations in mitochondrial genes in a lesser known Mizo population. The prevalence of T2D rise with family history of diabetes and clinical representation depends mostly on the severity of insulinopenia, lack of physical activity, obesity, demographic factor and involvement of genetic factors [14]. From this study, the risk for T2DM was found to be higher in patients with high smoked meat consumption followed by excess smoking and alcoholism. Saum, which is fermented pork fat, is one of the favorite foods of the Mizo people and is rich in hydrogenated oils. We also assessed the correlation between mtDNA gene mutations withrecognized prognostic relevance. Mitochondria play an imperative role in glucose metabolism, insulin secretion and biogenesis, hence its dysfunction is reportedly found to play a crucial role in diabetes development [2, 15]. Earlier reports showed the association of mitochondrial DNA mutations like 1310C < T, 1382A < C, 1438G < A, 1201A < G, 3243A < G, 3252A < G, 3256A < T, 3264A < C, 3271A < C, 3290T < C, 3303C < T, 3316G < A, 3394T < C, 8296A < G, 8344A < G, 11778G < A, 12026A <

Table 1 Polymorphisms in ATPase and ND1 genes of diabetic samples in Mizo population

Gene name	Frequency of mutation[a] (%)	Reference nucleotide	Nucleotide change	Nomenclature of mutation	Codon number	Codon position	Syn/Non syn	Codon change	Amino acid substitution	Reported/Disease caused	Polyphen2 score
ATP8	9.09	C	C > T	8414 C > T	17	1	Non-Syn	CTC > TTC	L > F	Reported for prostate cancer. Not for diabetes	0.99 (probably damaging)
ATP6	18.18	G	G > A	8584 G > A	20	1	Non-Syn	GCA > ACA	A > T		0.073 (Benign)
	9.09	T	T > C	8602 T > C	26	1	Non-syn	TTT > CTT	F > L		0.015 (Benign)
	9.09	G	G > A	8616 G > A	30	3	Syn	TTG > TTA	L > L	Reported for normal variation	—
	18.18	A	A > G	8701 A > G	59	1	Non-syn	ACC > GCC	T > A	Reported for diabetes	0.002 (Benign)
	9.09	G	G > A	8790 G > A	88	3	Syn	CTG > CTA	L > L	Reported for breast cancer. Not for Diabetes	—
ND1	9.09	G	G > A	3316 G > A	4	1	Non-syn	GCC > ACC	A > T	Reported for diabetes	0.00 (Benign)
	9.09	T	T > C	3394 T > C	30	1	Non-syn	TAT > CAT	Y > H		0.021 (Benign)
	9.09	T	T > A	3552 T > A	82	3	Syn	GCT > GCA	A > A		—
	36.36	C	C/T > T/C	3970 C > T	222	1	Syn	CTA > TTA	L > L		—
	9.09	A	A > G	4065 A > G	253	3	Syn	GAA > GAG	E > E		—
	27.27	C	C/T > T/C	4149 C > T	281	3	Syn	CGC > CGT	R > R	Reported for breast cancer	—

[a]Mutation frequency was calculated based on the total number of mutations obtained against the total number of cases

Table 2 Characteristic features of healthy and diabetic samples

Gene name	Nucleotide frequency (GC) %		Transition/Transversion bias (R)		Π		Θ		S		Ps		D	
	H	D	H	D	H	D	H	D	H	D	H	D	H	D
ATP6	44.3	44.4	132.25	151.55	0.0015	0.0021	0.0016	0.0025	11	2	0.0016	0.0073	−0.7099	−0.6484
ATP8	39.6	39.6	0.39	293.17	0.0000	0.0009	0.0000	0.0016	0	1	0.0000	0.0048	n/c	1.1285
ND1	47.7	47.7	0	2.23	0.0007	0.0017	0.0006	0.0015	1	4	0.0010	0.0042	1.6329	−0.3849

Abbreviations: *S* number of segregating sites, *ps* S/n, *Θ* ps/a1, *π* nucleotide diversity, *D* Tajima test statistic, *H* healthy, *D* diabetic

G, 12258C < A, 14577T < C, 14709T < C and 16189T < C [16–20] with T2D development. Particularly, mutation in tRNA Leu gene at 3243 (A < G) position and in the subunits of NADH dehydrogenase 1 and 4 have been reported to have strong association with incidence of diabetes in different populations [18, 19, 21, 22]. There was a significant correlation between the number of somatic mtDNA ATP6 mutations and the smoking and consuming betel-nut with paan (OR: 3.52; 95 % CI: 0.96–12.11) for the type 2 diabetes along with drinking alcohol (OR: 4.62; 95 % CI: 1.82–14.53). Besides, there was no significant difference in the concentration of plasma glucose level and familial history among the diabetic patients. Epidemic evidence have suggested that chronic smokers or tobacco consumers have a higher risk to be insulin resistant and exhibiting several aspects of the insulin resistance syndrome leading to the development of T2DM [23]. Earlier studies have identified a point mutation in the mitochondrial gene in a family with slowly progressive insulin-dependent diabetes mellitus (IDDM) or insulin-deficient non-IDDM. They have identified A to G transition at 3243 occurring in a highly conserved region of the tRNALeu (UUR) gene and this SNP and diabetes mellitus are maternally inherited and co-segregated [24–26].

High risk diabetic factors were seen in people with old age group belonging to low economic status with excessive meat intake and mostly prevalent among men. Obesity may also play an important role in triggering T2DM. History of familial inheritance is rarely seen in the case of Mizo Population. The Prevalence of Diabetes in India Study (PODIS) was carried out in 108 centres (49 urban and 59 rural) in different parts of India to look at the urban–rural differences in type 2 diabetes and glucose intolerance in the year 2004 [27, 28]. Our report is the first mitochondrial genetic alterations in diabetes reported from Mizo-Mongloid population.

Our results also revealed that non-synonymous variations were more frequent in the ATPase than in ND1 region of diabetic patients. ATP 6 belonging to ATP gene family is more mutated in this case than ATP 8. This indicates that simple sampling of blood would be advantageous for early marker development. The studied genes undergoes transitional substitution rather than transversion. Moreover, Tajima's D statistical testshows that ATP6mtDNA gene evolves randomly and NDI gene is evolving under a non-random process. This might depend on the directional selection or balancing selection or demographic expansion for Mizo population.

In our study, one major limitation is the small sample size which resulted in unstable risk estimates with wide 95 % CIs. The rate and standard deviation of mutation frequencies decreased with increasing sample size. There is a point beyond which increased sampling will have little impact on the accuracy and precision of estimates of mutation frequency. The risk estimates for ORs of diabetes in relation to lifestyle factors might have been biased, due to a small sample size and other factors such as selection bias. Another limitation of our study is that it does not explain the mitochondrial maternal inheritance of the mutations, because this study does not contain any familial sample.

Conclusion

To our knowledge, the present study is a novel finding in terms of the possible role of mtDNA ATPase and ND1 mutations in T2DM. ATP6 can be a good marker for the early detection of type 2 diabetes in Mizo-Mongoloid population. The mitochondrial gene alterations may attribute for diabetes risk along with the demographic habits and diet in Mizoram, Northeast Indian population. Besides clinical inconsistency, socio-economic status and environmental information needs to be considered in the assessment of risk profile of diabetic patients by health service.

Ethics, consent and permissions

The ethical committee of all institutes (Civil Hospital and Mizoram University, Aizawl, Mizoram, India) approved the study protocol involved in the study. All volunteers were fully informed about the study and participated with their full consent.

Additional file

Additional file 1: Table S1. Demographic and biochemical profiles of the diabetic patient samples from Mizo population. (DOC 106 kb)

Competing interests

The authors declare that they have no competing interests.

Authors' contributions

NSK, SG, JZ: conception and design of the study. FL, ST: Sample collection. FL, ST, SG: DNA isolation, PCR. FL, SG: Sequence and demographic analysis. FL, SG, NSK: drafting of the manuscript. JZ: critical revision of the manuscript for important intellectual content and supervision. All authors read and approved the final manuscript.

Acknowledgements

We thank all the blood donors for their voluntary support. This work is supported by the State Biotech Hub (BT/04/NE/2009) sponsored by the Department of Biotechnology (DBT), Govt. of India, New Delhi. We also thank Mr. David K. Zorinsanga, Department of Biotechnology, Mizoram University for his help during the work.

Author details

[1]Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India. [2]Department of Pathology, Civil Hospital, Aizawl 796001Mizoram, India.

Received: 12 November 2015 Accepted: 3 February 2016

References

1. Braunstein JB, Anderson GF, Gerstenblith G. Non cardiac comorbidity increases preventable hospitalizations and mortality among medicare beneficiaries with chronic heart failure. J Am Coll Cardiol. 2003;42:1226–33.
2. Gerbitz K, Gempel K, Brdiczker D. Perspectives in diabetes-mitochondria and diabetes, genetic, biochemical and clinical implications of the cellular energy circuit. Diabetes. 1996;45:113–26.
3. Rotiq A, Bonnefont JP, Munnich A. Mitochondrial diabetes mellitus. Diabetes Metab. 1996;22:291–8.
4. Elango S, Venugopal S, Thangaraj K, Viswanadha VP. Novel mutations in ATPase, ND1 and ND5 genes associated with peripheral neuropathy of diabetes. Diabetes Res Clin Pract. 2014;103:e49–52.
5. Morino K, Petersen KF, Dufour S, Befroy D, Frattini J, Shatzkes N, et al. Reduced mitochondrial density and increased IRS-1 serine phosphorylation in muscle of insulin-resistant offspring of type 2 diabetic parents. J Clin Invest. 2005;115:3587–93.
6. Heilbronn LK, Gan SK, Turner N, Campbell LV, Chisholm DJ. Markers of mitochondrial biogenesis and metabolism are lower in overweight and obese insulin-resistant subjects. J Clin Endocrinol Metab. 2007;92:1467–73.
7. Lamson DW, Plaza SM. Mitochondrial factors in the pathogenesis of diabetes: a hypothesis for treatment. Altern Med Rev. 2002;7(2):94–111.
8. Chen J, Hattori Y, Nakajima K, Eizawa T, Ehara T, Koyama M. Mitochondrial complex I activity is significantly decreased in a patient with maternally inherited type 2 diabetes mellitus and hypertrophic cardiomyopathy associated with mitochondrial DNA C3310T mutation: a cybrid study. Diabetes Res Clin Pract. 2006;74(2):148–53.
9. Howarth CC, Worsley A. Dietary habits of elderlypersons with diabetics. J Am Diet Assoc. 1991;91:553–7.
10. Dhillon PK. Breast cancer factsheet, South Asia network for chronic disease. Public health foundation of India. 2009. p. 1–22.
11. Lalmuanpuii R, Ghatak S, Pautu JL, Lallawmzuali D, Muthukumaran RB, Senthil-Kumar N. Mutation profiling in mitochondrial D-Loop associated with stomach cancer and tobacco consumers. J Clin Med Genom. 2015;3:1–5.
12. Ghatak S, Muthukumaran RB, Nachimuthu SK. A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech. 2013;24(4):224–31.
13. Xia X, Xie Z. DAMBE: software package for data analysis in molecular biology and evolution. Am Genet Assoc. 2001;4:371–3.
14. Devi K, Santhini E, Ramanan D, Ishwarya R, Prabhu NM. Mitochondrial ND1 gene mutation analysis in type II diabetes of Karaikudi population. Gene Genom. 2015;38:37–43. doi:10.1007/s13258-015-0337-7.
15. Choo Kang AT, Lynn S, Taylor GA, Daly ME, Sihota SS, Wardell TM, et al. Defining the importance of mitochondrial gene defects in maternally inherited diabetes by sequencing the entire mitochondrial genome. Diabetes. 2002;51:2317–20.
16. Liu SM, Zhou X, Zheng F, Li X, Liu F, Zhang HM, et al. Novel mutations found in mitochondrial diabetes in Chinese Han population. Diabetes Res Clin Pract. 2007;76:425–35.
17. Mezghani N, Mkaouar RE, Mnif M, Charfi N, Rekik N, Youssef S, et al. The heteroplasmic m.14709T < C mutation in the tRNAGlu gene in two Tunisian families withmitochondrial diabetes. J Diabetes Complicat. 2010;24:270–7.
18. Vijaya PV, Anitha S, Santhini E, Pradeepa D, Tresa D, Ganesan P, et al. Mitochondrial and nuclear gene mutations in the type 2 diabetes patients of Coimbatore population. Mol Cell Biochem. 2010;345:223–9.
19. Duraisamy P, Santhini E, Vijaya Padma V, Balamurugan R. Prevalence of mitochondrial tRNA gene mutations and their association with specific clinical phenotypes in type-II diabetes mellitus patients of Coimbatore. Genet Test Mol Biomark. 2010;14:49–55.
20. Weng SW, Lin TK, Wang PW, Chen SD, Chuang YC, Liou CW. Single nucleotide polymorphisms in the mitochondrial control region are associated with metabolic phenotypes and oxidative stress. Gene. 2013;531:370–6.
21. Kalinin VN, Schmidt W, Poller W, Olek K. A new point mutation in the mitochondrial gene ND1, detected in a patient with type II diabetes. Genetika. 1995;31:1180–2.
22. Zhao Q, Zhang L. Relationship between mitochondrial DNA nt3394T < C mutation in the ND1 gene region and DM. Chin J Med Genet. 2001;18:229–30.
23. Willi C, Bodenmann P, Ghali WA, Faris PD, Cornuz J. Active smoking and the risk of type 2 diabetes: a systematic review and meta-analysis. JAMA. 2007;298:2654–64.
24. Kadowaki H, Tobe K, Mori Y, Sakura H, Sakuta R, Nonaka I, et al. Mitochondrial gene mutation and insulin-deficient type of diabetes mellitus. Lancet. 1993;341:893–4.
25. Awata T, Matsumoto T, Iwamoto Y, Matsuda A, Kuzuya T, Saito T. Japanese case of diabetes mellitus and deafness with mutation in mitochondrial tRNALc"CUUR) gene. Lancet. 1993;341:1291–2.
26. Toledo FG. Effects of physical activity and weight loss on skeletal muscle mitochondria and relationship with glucose control in type 2 diabetes. Diabetes. 2007;56:2142–7.
27. Sadikot SM, Nigam A, Das S. The burden of diabetes and impaired glucose tolerance in India using the WHO 1999 criteria: prevalence of diabetes in India study (PODIS). Diabetes Res Clin Pract. 2004;66:301–7.
28. Sadikot SM, Nigam A, Das S, Bajaj S, Zargar AH, Prasanna-kumar KM. Diabetes India. The burden of diabetes and impaired fasting glucose in India using the ADA 1997 criteria: prevalence of diabetes in India study (PODIS). Diabetes Res Clin Pract. 2004;66:293–300.

Environmental Science and Pollution Research
https://doi.org/10.1007/s11356-018-3080-1

RESEARCH ARTICLE

Lifestyle chemical carcinogens associated with mutations in cell cycle regulatory genes increases the susceptibility to gastric cancer risk

Ravi Prakash Yadav[1] · Souvik Ghatak[1] · Payel Chakraborty[1] · Freda Lalrohlui[1] · Ravi Kannan[2] · Rajeev Kumar[2] · Jeremy L. Pautu[3] · John Zomingthanga[4] · Saia Chenkual[5] · Rajendra Muthukumaran[6] · Nachimuthu Senthil Kumar[1]

Received: 16 February 2018 / Accepted: 27 August 2018

Abstract

In the present study, we correlated the various lifestyle habits and their associated mutations in cell cycle (*P21* and *MDM2*) and DNA damage repair (*MLH1*) genes to investigate their role in gastric cancer (GC). Multifactor dimensionality reduction (MDR) analysis revealed the two-factor model of oral snuff and smoked meat as the significant model for GC risk. The interaction analysis between identified mutations and the significant demographic factors predicted that oral snuff is significantly associated with *P21* 3′UTR mutations. A total of five mutations in *P21* gene, including three novel mutations in intron 2 (36651738G > A, 36651804A > T, 36651825G > T), were identified. In *MLH1* gene, two variants were identified viz. one in exon 8 (37053568A > G; 219I > V) and a novel 37088831C > G in intron 16. Flow cytometric analysis predicted DNA aneuploidy in 07 (17.5%) and diploidy in 33 (82.5%) tumor samples. The G2/M phase was significantly arrested in aneuploid gastric tumor samples whereas high S-phase fraction was observed in all the gastric tumor samples. This study demonstrated that environmental chemical carcinogens along with alteration in cell cycle regulatory (*P21*) and mismatch repair (*MLH1*) genes may be stimulating the susceptibility of GC by altering the DNA content level abnormally in tumors in the Mizo ethic population.

Keywords Chemical carcinogens · Gastric cancer · Flow cytometer · Cell cycle · Mutation · Mizo population

Introduction

Cooking over an open flame is an ancient practice, starting from the early days of evolution of Homo sapiens. The consumption of grilled and smoked meat/vegetables seems to have increased in the modern food habits and population-wide changes in dietary style have dramatically increased stomach cancer rates in different parts of the world. There are so many delicacies prepared by smoking the food. The *N*-nitroso polycyclic aromatic hydrocarbons (NPAHs) are generated in the grilled and smoked meat/vegetables which will be the most harmful carcinogen causing stomach cancer (Correa et al. 1985; Ghatak et al. 2016). During the burning of wood by smoking and direct heat drying process, many harmful chemicals are formed, such as formaldehyde, polycyclic aromatic hydrocarbons (PAH), nitrogen and sulfur oxides, dioxins, heavy metals, etc. The PAHs are well known to cause several types of cancer in lab animals, such as liver, skin, and stomach (Stepanov et al. 2005, 2010).

Gastric cancer (GC) is the fifth most common cancer and is the third leading cause of cancer-related death worldwide

Ravi Prakash Yadav and Souvik Ghatak contributed equally to this work.

Responsible editor: Philippe Garrigues

Electronic supplementary material The online version of this article (https://doi.org/10.1007/s11356-018-3080-1) contains supplementary material, which is available to authorized users.

✉ Nachimuthu Senthil Kumar
nskmzu@gmail.com

[1] Department of Biotechnology, Mizoram University, Aizawl, Mizoram 796004, India

[2] Cachar Cancer Hospital and Research Centre, Silchar, Assam 788015, India

[3] Mizoram State Cancer Institute, Zemabawk, Aizawl, Mizoram 796017, India

[4] Department of Pathology, Civil Hospital, Aizawl, Mizoram 796001, India

[5] Department of Surgery, Civil Hospital, Aizawl, Mizoram 796001, India

[6] Department of Chemistry, Mizoram University, Aizawl, Mizoram 796004, India

Published online: 12 September 2018

(Ferlay et al. 2015). Globally, incidence of GC shows a wide geographic and ethnic variation, being particularly high in East Asian countries (Torre et al. 2015). In India, there are high GC incidence regions like Mizoram with an age-adjusted rate (AAR) of 50.6 and 23.3 per 10^5 populations in male and female, respectively (NCRP 2013). The general population in Mizoram is socially and ethnically unique from any other tribes and groups of India. Mizoram comprises of a distinct ethnic population with peculiar dietary habits such as consumption of smoked meat/vegetables, sa-um (fermented pork fat), nitroso salts, tobacco products; a tobacco smoke-saturated aqueous concentrate—tuibur, hand-rolled locally made cigarette—meiziol, freshly cut areca nut, slaked lime with half-betel leaf—kuhva, oral snuff and betel quid ("fresh" areca nut, slaked lime, condiments and coarse tobacco/pan masala wrapped in betel leaf for chewing), and alcohol (Phukan et al. 2006). A unique addiction of use of "tuibur" (tobacco smoke-infused aqueous answer) has been noticed in Mizoram (Madathil et al. 2018). Moist snuff is used by placing it between the lower lip or cheek and gum, and the nicotine in the snuff is absorbed through the tissues of the mouth. Moist snuff also comes in small, teabag-like pouches or sachets that can be placed between the cheek and gum. These are designed to be both "smoke-free" and "spit-free" and are marketed as a discreet way to use tobacco (Madathil et al. 2018). In Mizoram, manufactured smokeless tobacco products which are preferred as tobacco product packed in tear packs (gutka, khaini) and handmade cottage product packed in plastic packets (coarse tobacco mixed with slaked lime enriched water and trace levels of molasses known locally as "sahdah"). Hospital-based data from Mizoram have shown GC to be the most common cancer accounting for 30% of all cancer cases (Phukan et al. 2004). Hence, tobacco consumption may correlate with the high incidence of GC in Mizoram.

Gastric tumorigenesis is a multistep and multifactorial process associated with various genetic and epigenetic alterations including the activation of various oncogenes and inactivation of tumor suppressor genes and mismatch repair (*MMR*) genes (Igaki et al. 1994). According to multistep model of gastric carcinogenesis, the most common and principle pathway affected is cell cycle by genetic aberrations. Cell cycle progression is a highly ordered biological process; hence, alterations in the cell cycle genes has been suggested to contribute the underlying the tumorigenesis of GC (Decesse et al. 2001).

P21 has been reported to play multiple roles within the cell including cell cycle regulation, senescence, apoptosis, DNA repair, and differentiation (Parker et al. 1995; Ciccarelli et al. 2005; Jung et al. 2010). Although mutations in *P21* are infrequent in human cancers (Shiohara et al. 1994), previous studies have shown that *P21* may act to either promote or suppress in various cancers. *P21* is a putative tumor suppressor gene and its mutations have been studied as a risk factor in various cancers (Watanabe et al. 1995), including GC (Mousses et al. 1995; Bahl et al. 2000). The murine double minute 2 (*MDM2*) is one of the central nodes in the p53 pathway and can control p53 protein levels and activity. *MDM2* gene encodes an important negative regulating protein which promotes ubiquitin-dependent proteosomal degradation of p53 by functioning as an E3 ubiquitin ligase (Oren et al. 2002; Bouska et al. 2008). The mismatch repair (*MMR*) system plays an essential role in identifying and rectifying replication errors as well as additional errors in DNA which may arise through physical or chemical damage. The pathogenic alterations are scattered in the carboxyterminus domain of MLH1 protein and the position annotated as pms2, mlh3, and pms1 interaction domain (Guerrette et al. 1999; Lipkin et al. 2000; Kondo et al. 2001).

In Mizo tribal population, there is limited evidence for the genetic and environmental risk factors that may be associated with stomach cancer (Ihsan et al. 2011; Malakar et al. 2012). In the present study, a case–control study for the high prevalence of GC in Mizoram has been attempted in order to identify the mutations in cell cycle genes, *P21* and *MDM2*, besides DNA damage repair gene, *MLH1*. Further, the correlation between these mutations and the environmental as well as dietary factors that seem to play an important role in GC etiology are also described.

Material and methods

Subjects

This study included a cohort of patients with pathologically confirmed gastric tumor. A total of 40 gastric tumor tissues (28 males and 12 females) and their matched adjacent normal gastric mucosa were collected. Peripheral blood samples were also collected from the patients and sex-age matched healthy individuals. All the gastric tumor and adjacent normal samples were collected from Civil Hospital and Genesis Laboratory and Diagnostics, Aizawl, Mizoram. The demographic information such as age, gender, dietary habits, familial incidence of cancer, smoking habits, and alcohol consumption were obtained after getting informed consent using a structured questionnaire. All the study participants received written information and gave consent for the publication of the required results. Ethics approval for this study was obtained from the Institutional Review Board (IRB) of the Civil Hospital, Aizawl (B.12018/1/13-CH(A)/IEC). Hematoxylin and eosin-stained slides were prepared from paraffin block of tumor tissues after micro-dissection to determine the type gastric adenocarcinoma and TNM staging.

DNA extraction and PCR amplification

Genomic DNA was isolated from the blood following the standard protocol (Ghatak et al. 2013). The genomic DNA was extracted from the paraffin embedded tumor tissue and adjacent normal by the modified protocol of Ghatak et al. (2014). All exons and adjacent intronic regions of the *P21*, *MDM2*, and *MLH1* genes were screened. PCR (Eppendorf, USA) was carried out in 25 μl total reaction volumes (containing 100 ng template DNA, 0.2 pM of each primer, 1× PCR buffer, 1.5 mM $MgCl_2$, 200 mM dNTPs, 1 unit Taq DNA polymerase (Fermentas Inc., Glen Burnie, MD). The reaction mixture was heated to 94 °C for 7 min, followed by 40 cycles, each consisting of 1 min denaturation at 94 °C, 1 min annealing at 63 °C, 1 min extension at 72 °C, and a final 7 min extension at 72 °C (Supplementary Table 1). The PCR amplification products (3 μl) were subjected to electrophoresis on 1.2% agarose gel in 1× Tris-acetate-EDTA buffer at 80 V for 30 min and stained with ethidium bromide (Himedia, India) and images were obtained in gel documentation (G-Box; Syngene, UK) system (Fig. 1).

Sequencing and sequence analysis

Sequencing of the exons of these genes is most useful when using primers that include a portion of the intron/exon boundary. This allows the entire exon to be sequenced as well as the splice sites where mutations are known to occur (Boland and Goel 2010). The impact of amino acid allelic variants on protein structure/function can be predicted after performing multiple sequence alignments and protein 3D structures. The Sorting Intolerant from Tolerant (SIFT) algorithm was applied. SIFT is a program that predicts the effect of amino acid substitutions on protein function, on the basis of sequence conservation during evolution and the nature of the amino acids substituted in a gene of interest. SIFT scores were calculated online (http://sift.jcvi.org/). If the value is less than 0.05, the amino acid substitution was predicted as intolerant, while those with a value greater than or equal to 0.05 were classified as tolerated. Human Splicing Finder (http://www.umd.be/HSF/) was used for finding the splice site region.

Splicing donor/acceptor sites and branch point sequences

To predict potential 5′ss and 3′ss (splicing site), we used matrices-derived splicing finder database. A potential splice site is defined as an n-mer sequence. For each "n" position, a weight is given to each nucleotide, based on its frequency and the relative importance of its position in the sequence motif (position weight matrices, PWM). Only n-mer sequences with consensus values (CV) higher or equal to a given threshold are considered as potential 5′ or 3′ss. The human branch point (BP) consensus sequence is YNYCRAY, and the threshold for BP sequences was fixed at 67. Since many intronic sequences match the BP consensus sequence, hence, we

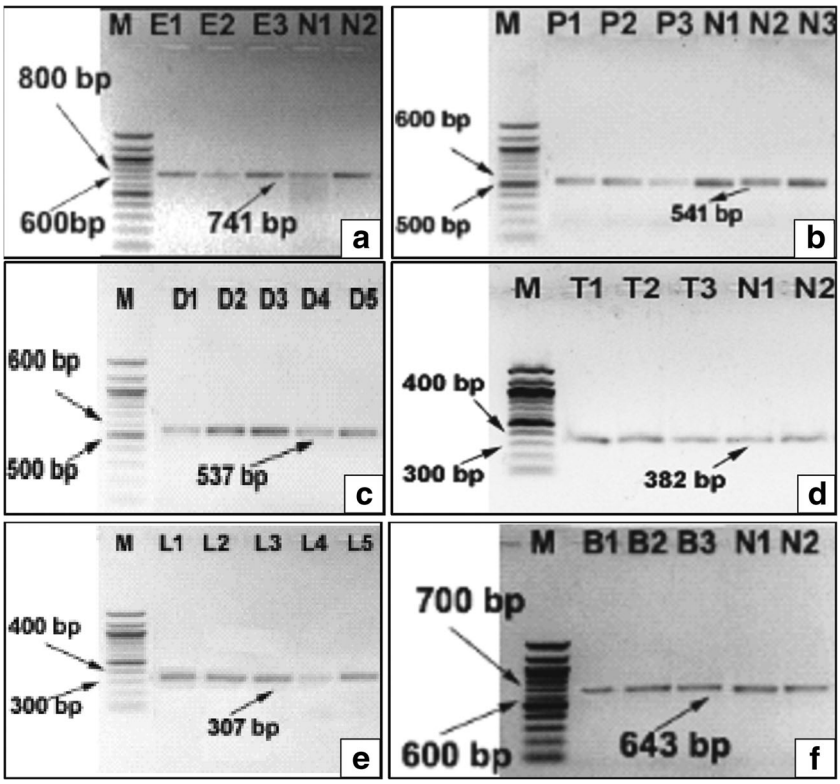


Fig. 1 PCR products of mdm2 gene. **a** Exon 2, **b** Exon 8. Lanes M: 100 bp DNA ladder (Invitrogen, USA), D1, D2, D3, D4, D5 and T1, T2, T3 = GC samples, N1, N2 = controls. Gel images of the PCR products of p21gene. **c** Exon 2, **d** Exon 3. Lanes M: 100 bp DNA ladder (Invitrogen, USA), E1, E2, E3 and P1, P2, P3 = GC samples, N1, N2, N3 = controls. Gel images of the PCR products of mlh1 gene. **e** Exon 8, **f** Exon 16. Lanes M: 100 bp DNA ladder (Invitrogen, USA), L1, L2, L3, L4, L5 and B1, B2, B3 = GC samples, N1, N2 = controls

included the AG-Exclusion Zone algorithm (Gooding et al. 2006) to predict BP candidates. Splicing donor/acceptor sites and BP was splice site finding was estimated by Human Splicing Finder (http://www.umd.be/HSF/). HSF searches all AG dinucleotides that are included in a 3′ss candidate sequence (threshold of 67) and therefore define the exclusion zones for a given intronic sequence and its intron–exon boundary. HSF annotates the functional BP as the strongest candidate without a 3′-exclusion zone before the natural 3′ss because it has been shown that the BP allows the recognition of the first downstream 3′ss. Additionally, to take into account the steric obstruction caused by the spliceosome, we excluded BP sequences located at less than 12 nt from the exon. Finally, as most BP sequences are located between − 21 and − 34 nt from the exon and only a window of 100 bp is processed.

DNA content analysis

Flow cytometry measures DNA contents (ploidy) of cancer cells and rate of proliferation, indicating the proportion of cells under DNA synthesis (S-phase fraction) and has been shown to yield prognostic information in many of the human malignancies (Merkel and Mcguire 1990). S-phase fraction (SPF) is also a well-known independent prognostic factor in some human malignancies, such as breast, prostate cancer, and gynecological malignancies (David and Hedley 1994).

For the analysis of cell cycle distribution, paraffin-embedded tumor tissue was used following modified method described by David and Hedley (1994); 10^6 cells were harvested by centrifugation, washed in phosphate-buffered saline (PBS) (Sigma P4170), fixed with ice cold 70% ethanol, and treated with 1 mg/ml RNAse for 30 min. Intracellular DNA was labeled with propidium iodide (50 mg/ml) and incubated at 4 °C in the dark. Samples were then analyzed using flow cytometer FACSCalibur (BD, Germany). The data obtained was analyzed using the ModFit LT software (DNA Modeling System) version 2.0 (Verity Software House, Inc.) and single parameter histograms was obtained.

Multifactor dimensionality reduction analysis

The multifactor dimensionality reduction (MDR) is a nonparametric, genetic model-free statistical approach to identify high-order gene–gene and gene–environment interactions associated with GC risk (Hahn et al. 2003; Cattaert et al. 2011). It is applied for overcoming the sample size limitations. In the present study, MDR software package (MDR 3.0.2) was used to generate the best one-dimensional multifactor model to classify and predict GC susceptibility. The best model was selected based on maximum cross-validation consistency (CVC) and testing balance accuracy (TBA). The MDR permutation results were considered to be statistically significant at the 0.05 level (Ritchie et al. 2003; Manuguerra et al. 2007).

Interaction entropy graph

Interaction graphs were built to visualize and interpret the results obtained from MDR. Entropy estimates were used to determine the information gain about a class variable (e.g., case–control status) from merging two variables together. Entropy estimates are useful for building interaction graphs facilitating the interpretation of relationships between variables (Choudhury et al. 2015).

Statistical analysis

The polymorphism and demographic factor in each group were estimated for their association with GC using odds ratios (ORs) and 95% confidence intervals (CIs) in the logistic regression (LR) model adjusted with multivariable analysis. Each polymorphism was checked by the presence and absence of the SNPs. Additionally, logistic regression analyses were conducted to compute the influence of both genetic and environmental factors for GC. For all tests, a two-sided p value < 0.05 was considered statistically significant. All statistical analyses were performed using SPSS 20.0 program (SPSS Ibérica, Madrid, Spain) and SYSTAT 13.0. (Systat Software Inc., USA). Hardy–Weinberg equilibrium by a chi-square (χ^2) test with one degree of freedom (df) was performed between case and control subjects. Fisher's exact test also used for comparing the demographic and habits between patients and controls. Correlation between clinicopathological features and DNA content was estimated by SPSS 20.0 program.

Results

Characteristics of study subjects

The frequency distributions and selected characteristics of the patients and controls are presented in Table 1. The median age was 58.7 years for the patients and 52.18 years for the controls. The analysis indicates that smoked meat/vegetable (OR 16.214; 95% CI 2.746–95.749; p 0.002) and oral snuff (OR 10.496; 95% CI 2.410–45.710; p 0.002) are the major risk factor for GC among our study population (Table 1).

MDR and interaction entropy analysis

MDR analysis was performed to explore the potential gene–gene-environment interaction. In the present study for the entire dataset, smoked meat is the best one-factor model found statistically significant ($p < 0.0001$) with a CVC of 8/10 and testing accuracy of 0.6653. The combination of smoked meat and oral

Table 1 Demographic characteristics of the cases and control samples

Demographic factor	HC ($n = 40$)	GC ($n = 40$)	ORs (95% CI)	p value
Age years ± SD (range)	52.18 ± 12.35	58.7 ± 9.76	–	
Male	12 (30%)	28 (70%)	–	
Female	28 (70%)	12 (30%)		
Sa-um	29 (72.5%)	37 (92.5%)	0.979 (0.346–2.770)	0.968
High salt intake	30 (75%)	31 (77.5%)	0.507 (0.077–3.340)	0.480
Smoked meat/vegetable	25 (62.5%)	38 (95%)	16.214 (2.746–95.749)	0.002
Pickle	22 (55%)	23 (57.5%)	0.340 (0.108–1.072)	0.065
Tuibur consumption	14 (35%)	17 (42.5%)	0.755 (0.350–1.631)	0.475
Cigarette smoking	24 (60%)	29 (72.5%)	2.091 (0.810–5.400)	0.127
Oral snuff	6 (15%)	20 (50%)	10.496 (2.410–45.710)	0.002
Tiranga/Gutkha	8 (20%)	4 (10%)	8.954 (0.816–98.308)	0.073
Kuhva(betel nut, slaked lime wrapped in betel leaf)	15 (37.5%)	17 (42.5%)	1.094 (0.394–3.036)	0.864
Family history of gastric cancer	5 (12.5%)	7 (17.5%)	2.148 (0.579–7.970)	0.253
Family history of other cancers	5 (12.5%)	3 (7.5%)	1.011 (0.369–2.769)	0.983

HC healthy control, *GC* gastric cancer, *OR* odd ratio, *95% CI* 95% confidence interval

snuff was found to be the best two-factor model which was also the best overall model with a CVC of 10/10 and TBA of 0.7389 ($p < 0.0001$). The combination of sa-um, smoked meat, and tuibur was found to the best three-factor model with a CVC of 3/10 and TBA of 0.4042 ($p < 0.0001$) (Fig. 2a). The previous statistical analysis results were reproduced in MDR analysis also.

Interaction entropy graphs were created using MDR results, for better verification and visualization of interactions between gene–environment factors. In interaction entropy graph, smoked meat showed the highest independent effect (20.56%) and also had moderate synergistic interaction with sa-um (0.43%). Oral snuff (8.19%) also explained considerable entropy independently (Fig. 2b).

The association between gene mutations and GC risk

A binary logistic regression model was applied to estimate the association between gene mutations and risk of GC (Table 2). The mutations in *P21* gene at 3′ UTR were associated with oral snuff consumption in GC patients (OR 9.256; 95% CI

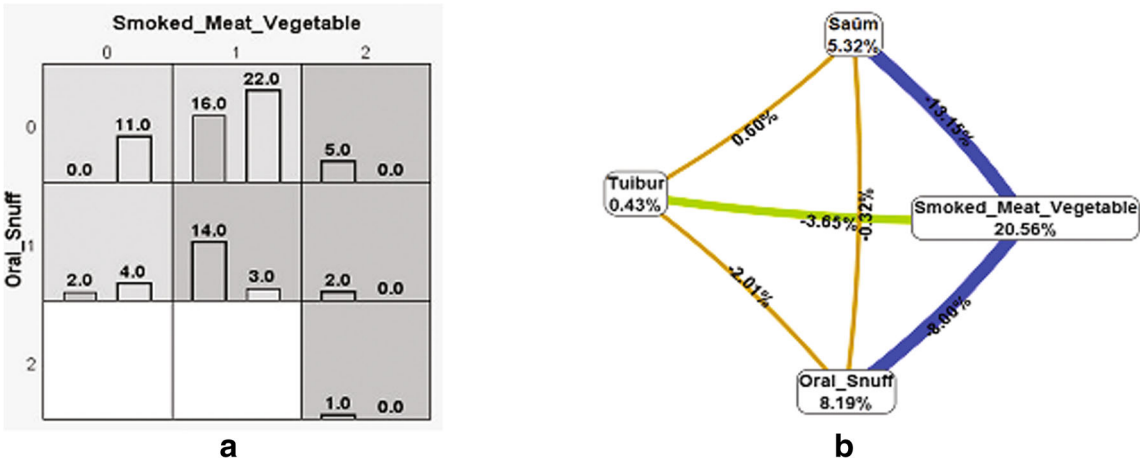


Fig. 2 Multifactor dimensionality reduction (MDR) analysis. **a** The summary of the two-factor model (smoked meat and oral snuff) predicted by MDR is represented in the graph. The distribution of high-risk (dark shading) and low-risk (light shading) combinations associated with GC risk. For smoked meat and oral snuff, 0 represents less consumption, 1 represents moderate consumption, and 2 represents high consumption. **b** Interaction entropy graph. The percent of the entropy for independent factors as well as their interactions are represented in the graph where positive percentage of entropy denotes synergistic interaction while negative percentage denotes redundancy. The red color indicates a high degree of synergistic interaction and orange a lesser degree, whereas gold represents midpoint and blue represents the highest level of redundancy followed by green

Table 2 Interaction between mutations and significant demographic factors

Factor	Gene name	Position	ORs (95% CI)	*p* value
Oral snuff	p21	Intron 2	1.025 (0.253–4.150)	0.972
		3′ UTR	9.256 (1.842–46.509)	0.007
	mlh1	Exon 8	0.956 (0.183–4.986)	0.958
		Intron 16	1.732 (0.143–20.956)	0.666
Smoked meat/vegetable	p21	Intron 2	4.149 (0.970–17.738)	0.050
		3′ UTR	0.728 (0.182–2.909)	0.653
	mlh1	Exon 8	1.510 (0.324–7.043)	0.600
		Intron 16	1.386 (0.109–17.543)	0.801

OR odds ratio, *95% CI* 95% confidence interval

1.842–46.509; $p < 0.007$). Similarly, smoked meat/vegetable and *P21* intron 2 (OR 4.149; 95% CI 0.970–17.738; $p < 0.050$) were also found to be significantly associated with increased risk for GC (Table 3). However, the other genes did not show any association with the demographic factors.

Sequence variations of *P21*, *MDM2*, and *MLH1* genes and their consequence

The molecular analysis revealed a total of five mutations in P21 gene (Table 2, Fig. 2). In intron 2, mutations 36651738G > A, 36651804A > T, and 36651825G > T were identified in 5% of total GC samples and were found to be novel (not reported previously in the database). The mutations were of single base change type. Splice site changes were identified as a result of 36651738G > A mutation, and it was also predicted that this might affect the protein folding and/or functional features. Splice site donor is marginally increased (wt 0.8042/mu 0.8719). No variation in potential splice site changes were identified due to 36651804A > T, whereas, 36651825G > T was found to increase marginally at the splice site donor (wt 0.8725/mu 0.9361). 3′UTR region of *P21* showed two known (previously reported) mutations

Table 3 Polymorphism and mutations in cell regulatory genes of gastric cancer samples

Gene name	Position	Nomenclature of mutation	Frequency of mutation (%)	AA change	PolyPhen-2/SIFT / PROVEAN	Novel/ reported	Effect of mutation by mutation taster
p21	Intron 2	36651738G > A	5	–	–	Novel	Polymorphism (single base change) Protein features (might be) affected Splice site changes (donor marginally increased wt 0.8042/mu 0.8719)
		36651804A > T	5	–	–	Novel	Polymorphism (single base change) No abrogation of potential splice site
		36651825G > T	5	–	–	Novel	Polymorphism (single base change) Donor marginally increased (wt 0.8725/ mu 0.9361)
	3′UTR	36653580C > T	10	–	–	Reported in the database	Polymorphism (single base change) Splice site changes (splice site change occurs after stop codon, acceptor marginally increased, wt 0.53/ mu 0.64)
		36653597C > T	5	–	–	Reported in the database	Polymorphism (single base change) Splice site changes (splice site change occurs after stop codon, acceptor marginally increased, wt 0.5311/ mu 0.5459)
mlh1	Exon 8	37053568A > G	10	219I > V (ATC > GTC)	(0.018) Benign/ neutral/tolerated	Reported n the database	In the protein structure helix (212–220) might be lost Splice site changes (wt 0.7064/mu 0.7505, acceptor marginal increased)
	Intron 16	37088831C > G	5	–	–	Novel	Polymorphism (wt 0.5187/mu 0.5635, acceptor marginal change)

36653580C > T in 10% and 36653597C > T 5% of GC samples. These mutations are affecting the splice site change by acting after stop codon. Splice site acceptor is marginally increased in both the cases. Annotated sequences were deposited in EBI repository with accession number (LN997431-LN997630).

Whereas in *MLH1* gene, two variants were identified one in exon 8 (37053568A > G; 219I > V) and other in intron 16 (37088831C > G) of which intronic variant is novel (37088831C > G) (Table 2, Fig. 3). The SIFT score for the *MLH1* exonic variant ((37053568A > G; 219I > V)) demonstrated that it is tolerated and may elicit only minor effect on the protein structure. 37088831C > G variant in intron 16 of *MLH1* gene was identified in 5% of GC samples which is also novel (not reported in database) and found to affect splice site by increasing the acceptor marginally (wt 0.5187/mu 0.5635). In the present study, we expected to find common pathogenic mutations, but for this cohort, none were found in *MDM2* gene. This result is indicative of the fact that there may not be any common or founder mutations for *MDM2* gene in Mizo population. Prediction of structural variation between wild- and mutant-type amino acids for *MLH1* exonic variant ((37053568A > G; 219I > V)) was carried out by HOPE analysis which indicated that the mutant type is smaller in size than the wild type, affecting the intramolecular and external interactions due to clashes (Fig. 4).

DNA content analysis

Flow cytometric analysis (FCM) analysis can provide not only the kinetic estimates such as the fraction of cells in S-phase (SPF) but also capable of subdividing neoplasms into DNA diploid or DNA aneuploid tumors based on the presence of different sub-populations in different phases of cell cycle. FCM analysis predicted DNA aneuploid in 07 (17.5%) and diploid in 33 (82.5%) diploid tumor samples (Fig. 5). Significantly higher S-phase fraction (SPF) was observed in all the GC samples (51.24–72.09) compared to controls (32.45–44.12). The G2/M phase was found arrested in gastric tumor samples. The G2/M phase is found to be arrested in the gastric tumor samples besides more DNA content in S-phase (Table 4, Fig. 2).

Splicing abnormality due to mutations

The difference between wild-type (wt) active sites and mutant-type inactive sites was predicted by the HSF algorithm and was calculated by the consensus values (CV) of 50ss or 30ss. The CV higher than 80 represent as a stronger relation with active sites and between 70 to 80 represent a weaker relation with active site. Mutations can create a new cryptic splice site rather than disruption of a 50ss or a 30ss active sites which was correctly predicted by HSF. 36653580C > T mutation in 3′UTR of *P21* gene showing 1.53% CV variation between wild type (86.39) and mutant (84.86) type. Due to the potential CV change, the branch point motifs also potentially changed for wild type (CGCCCAC) and mutant type (TGCCCAC) (Table 5).

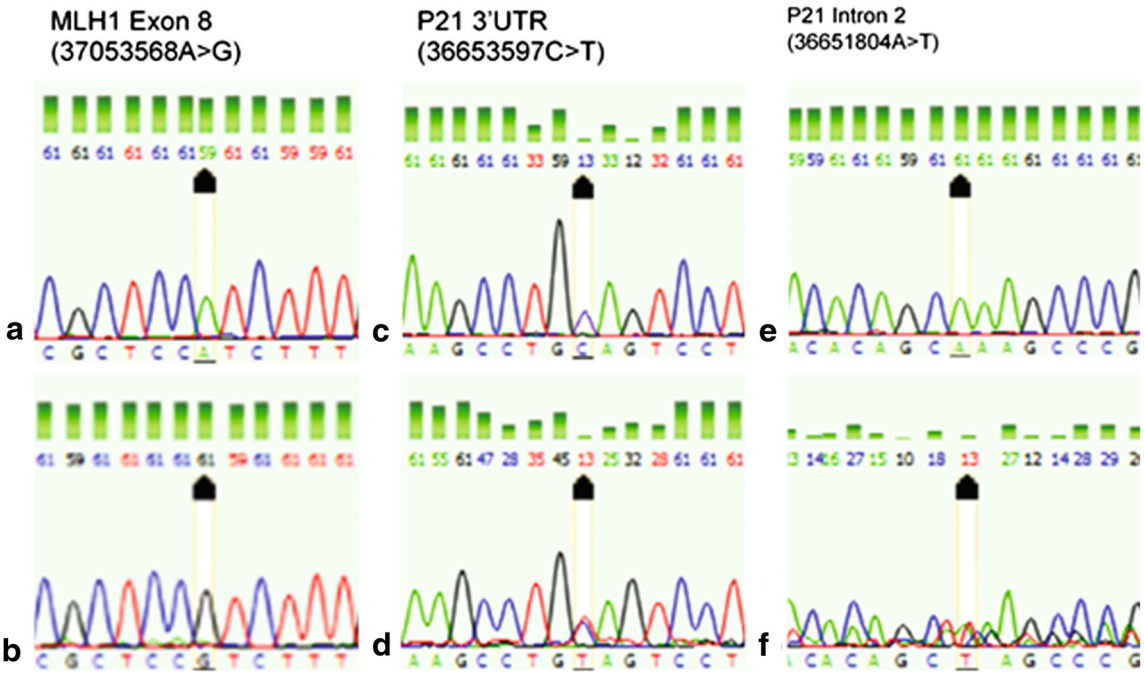


Fig. 3 Electropherogram of the genes from GC samples (**b**, **d**, **f**) compared with healthy samples (**a**, **c**, **e**)

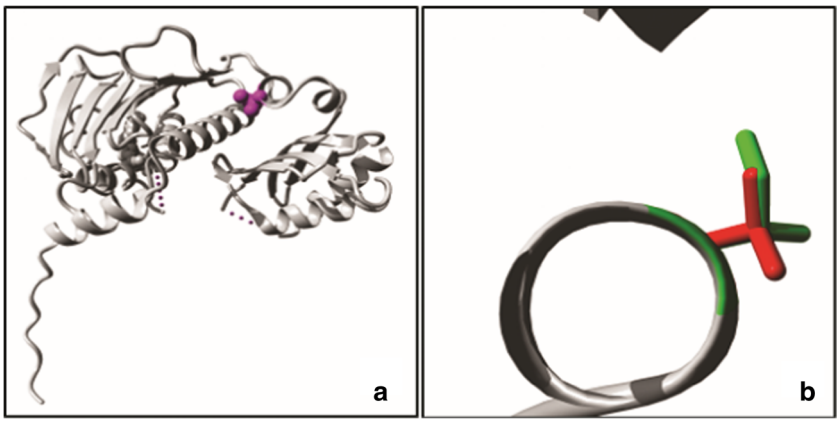


Fig. 4 3D structure of the mutation (37053568A > G; 219I > V) in mlh1 gene. **a** Complete chain of protein and the pink spheres and rest of the protein is shown in gray representing the site of mutation (219I > V). **b** Mutation of isoleucine to valine at position 219 due to 37053568A > G. Wild-type and mutant-type side chains are shown in green and red, respectively

Discussion

Our attempt was to accumulate evidence to identify the relationship between chemical carcinogens and GC risk in relation to gene mutations. GC has positive association with consumption of smoked and salted meat/vegetables (Correa et al. 1985; Ghatak et al. 2016). *N*-nitroso compounds can be generated in meat during smoke drying and preservation which are highly carcinogenic due to the reaction between nitrite with amines and amides, which is found in meat and other proteins (Correa et al. 1985; Ghatak et al. 2016). The nitrite present in smoked meat play a secondary role in the progress of chronic atrophic gastritis, which can develop as stomach cancer in the later stages (Nomura et al. 1990). Previous studies have reported positive association of high intake of smoked meat as potential confounder for GC (Kneller et al. 1992; Appelman et al. 1992; Ward and Lopez-Carrillo 1999). In the present study, smoked meat and high salt intake was positively associated with GC. Intra-gastric high salt accumulation causes the expansion of surface mucous prompting for aggravation and damage, for example, diffuse erosion, atrophic gastritis, and diminished corrosiveness of the stomach which creates a condition supporting *H. pylori* infection (Tsugane et al. 2004; Tsugane and Sasazuki 2007). Gastric mucus can also be damaged by smoked meat intake with extra salt, leading to increased epithelial cell proliferation as part of the repair process (Campos et al. 2006).

In Mizoram, mostly pork, beef, fresh water fish, birds, and/or animal meat along with the seasonal vegetables are used for heat drying (traditional wood-burning) for preservation. Higher concentrations of polycyclic heterocyclic amines (PAHs) formed during the preparation of food at higher temperature conditions such as frying, roasting, and/or grilling (Phillips 1999). Interestingly, only modest levels of PAHs are formed, while cooking the food by steaming/stewing/boiling. Along with lean meat, heat drying of "meat with fat" besides "skin and fat" of pork meat is preferred in Mizoram for preservation. Few of the PAHs and nitrosamines are carcinogenic, while some of the PAHs, HAAs, and volatile nitrosamines are indeed pro-carcinogens. These pro-carcinogenic species are metabolically activated, after being ingested and metabolized subsequently as carcinogens (Hecht 2003). In addition, soda (sodium

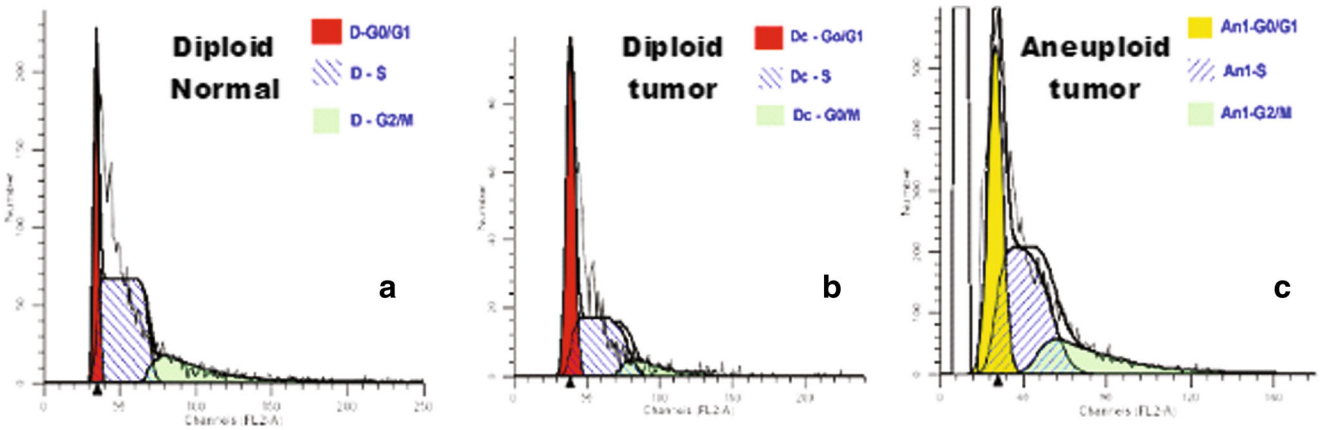


Fig. 5 DNA content analysis in gastric cancer samples measured by the flow cytometric profile. Histogram of **a** diploid normal sample with G0/G1 peak (red), S-phase (shaded peak) and G2/M peak (green). **b** Diploid gastric cancer tumor sample. **c** Aneuploid gastric cancer tumor sample with aneuploidy peak (yellow)

Table 4 Distribution of different phase fraction according to DNA content

Sample no.	Ploidy status	Ploidy (%)	G0/G1 phase			S-phase			G2/M phase	
			Low (%)	Moderate (%)	High (%)	Low (%)	Moderate (%)	High (%)	Low (%)	High (%)
Healthy (40)	Diploid	40 (100)	1 (2.5)	13 (32.5)	26 (65)	38 (95)	2 (5)	0	3 (7.5)	37 (92.5)
Cancer (40)	Diploid	33 (82.5)	18 (54.5)	15 (45.4)	0	0	3 (9.09)	30 (90.9)	28 (84.8)	5 (15.1)
	Aneuploid	7 (17.5)	0	1 (14.3)	6 (85.7)	0	1 (14.3)	6 (85.7)	7 (100)	0

bicarbonate)—an alkaline preparation, frequently used as food additives was significantly associated with increased risk of stomach cancer in Mizoram (Phukan et al. 2006).

During consumption of oral snuff and chewing tobacco, the harmful contaminant can mix with saliva and enter inside the stomach, although the association between oral snuff and stomach cancer incidence was negatively reported for some study and one study clearly reported the positive associations (Chao et al. 2002; Furberg et al. 2006). Oral snuff production in other parts of the world involves heat processing rather than curing and fermentation as is done in the Mizoram, northeast India. Fermentation of tobacco can generate higher amounts of tobacco-specific *N*-nitrosamines and volatile *N*-nitrosamines due to the high *N*-nitrosation of nicotine.

It is important to note that higher concentration levels of PAHs (Phillips 1999) and lower concentration levels of HAAs in pork meat with higher fat content was observed (Chen et al. 1990). PAHs such as benzo(a) pyrene formed in smoked food have been correlated in many areas of the world with high stomach cancer rates (Yeh et al. 2009). Furthermore, oral snuff and other tobacco products contain carcinogens like the nitrosamines which acts as cofactor for pathogenesis of GC (Curado et al. 2007). Gutkha, khaini, and sahdah demonstrably contains relatively high concentration levels of tobacco-specific nitrosamines (TSNAs) (some of which are proven carcinogens to human, while some of which are pro-carcinogens), nitrate, and/or nitrite as well as PAHs (Stepanov et al. 2005, 2010). It may be possible that PAHs besides nitrite and TSNAs may play an important role in the tumorigenesis of GC patients in Mizoram, due to the higher level of PAHs, nitrite, and TSNAs exposed in the lifestyle habits.

It is widely accepted that genetic and environmental factors are major etiological factors for GC. In the present study, we investigated the spectrum of mutations in genes *P21*, *MDM2*, and *MLH1* among GC patients. Among the genetic factors, cell cycle regulatory genes besides DNA mismatch repair gene were found to be associated with various cancers including GC (Gartel and Tyner 1999). Cell cycle control is crucial for the normal cell growth and differentiation. The present study also attempted to establish a potential association of alteration in these gene variations with demographic and dietary factors. In this study, we have tried to establish a significant relation between driver gene mutations and epidemiological factor for diffuse type GC. To the best of our knowledge, the present study results are the first report of the association of environmental factors with mutations in cell cycle regulatory genes suggesting the implication of genetic alterations and its correlation with cell cycle in GC development.

The high prevalence of GC in Mizoram has been attributed to peculiar dietary habits (viz. high consumption of smoked meat, salt-preserved foods, dietary nitrite, traditional fermented food, and heavy addiction to various tobacco products and alcohol (Phukan et al. 2005, 2006). The demographic factors play major role in pathogenesis of GC. A hospital-based matched case–control study showed an elevated risk of stomach cancer in case of frequent consumption of sa-um (fermented pork fat) and smoked dried salted meat and fish (Phukan et al. 2006). Smoking as a variable risk factor for

Table 5 Mutations leading to splicing defects

Gene	Mutation	Position	WT branch point motif	Mutant branch point motif	WT CV	Mutant CV	CV variation (%)
p21	Intron 2	36651738G > A	AAGCAGG	AAGCAAG	9.95	39.57	29.62
		36651804A > T	AGCAAG	AGCTAG	61.53	65.52	3.99
		36651825G > T	ATAGTGT	ATATTGT	6.23	21.31	15.08
	3′UTR	36653580C > T	CGCCCAC	TGCCCAC	86.39	84.86	− 1.53
		36653597C > T	CTGCAGT	CTGTAGT	16.53	25.56	9.03
mlh1	Exon 8	37053568A > G	CTCCATC	CTCCGTC	46.15	53.04	6.89
	Intron 16	37088831C > G	TTGACAG	TTGAGAG	47.68	42.74	− 4.94

A new site was created by the mutation; the motif was abolished by the mutation

stomach cancer has also been reported from India (Dikshit et al. 2012). In the present study, the results of demographic study indicated that males are more susceptible to GC than females. The putative association between the risk of GC and unique dietary habits has been controversial for decades. In this study, we found that high intake of smoked meat/vegetable (16.214 (2.746–95.749); $p = 0.002$) and oral snuff (10.496 (2.410–45.710); $p = 0.002$) are significant risk factors for the high incidence of GC (Table 1). However, the reason for disparities is still not yet known, although recent research has suggested that genetic factors may be the reason for differences in GC susceptibility of various populations (Yan et al. 2015). Also, family history of other cancers are found to be associated with the increased risk of GC in this population which may be because of their genetic make-up and inheritance of faulty genes which renders them predisposed to cancer (Yaghoobi et al. 2010). The Mizo population is mongoloid in origin and distinct from the rest of India in terms of their diet, lifestyle, and geographical distribution (Ghatak et al. 2013). Also, in other mongoloids like Japanese, daily consumption of meat among women was found to increase the risk of GC by 6.5-fold (Santarelli et al. 2008). In another study, the potential causal role of tobacco was observed in high-risk area of China, where smoking was found to nearly double the risk of transition to gastric dysplasia (Piazuelo and Correa 2013). In the present study, the interaction between identified mutations and the significant demographic factors, smoked meat/vegetable and oral snuff were found to be associated with risk of GC, with a significant association between oral snuff and *P21* 3′UTR mutations similarly association between smoked meat/vegetable and *P21* Intron 2 mutations. Consumption of oral snuff ($p = 0.002$) was significantly associated with GC followed by Tiranga/Gutkha consumptions ($p = 0.073$) (Table 3).

Cell cycle deregulation is common pathway in pathogenesis of human cancer, and alteration of *P21*, the cell cycle regulator, is involved in the development of many human malignancies (Gartel 2005; Lin et al. 2011).The molecular analysis revealed in total five mutations in P21 gene. In the intron 2 of *P21* gene, 36651738G > A, 36651804A > T, and 36651825G > T novel mutations were identified in 5% of total GC samples. Splice site changes were identified as a result of 36651738G > A and 36651825G > T mutation, and it was also predicted that this might affect the protein structural features as the splice site donor is marginally increased. The *P21* CDK inhibitor gene is located at 6q21.2, and its expression has been shown to be regulated largely at the transcriptional level by both p53-dependent and independent mechanisms by a variety of transcription factors that are induced by a number of different signaling pathways (An et al. 2014). Previous studies demonstrated that FOXA2, transcription factor activation of *P21* transcription via direct binding to the P21 promoter and affects the activity of *P21* gene, which results in cell cycle arrest at the G1 phase and inhibition of cell proliferation in p53-deficient cell (Wang et al. 2012). 3′UTR region of *P21* showed two known (previously reported) mutations 36653580C > T in 10% and 36653597C > T 5% GC samples. 36653580C > T polymorphism is thought to cause a functional change in *P21* due to generation of a cryptic spice site by acting after stop codon, and as this polymorphism lies in a crucial region for cell differentiation, proliferation may increase cancer risk by altering messenger RNA stability, which, in turn, may affect protein expression and activity (Campbell et al. 2009). Mutations or single nucleotide polymorphisms (SNPs) in the *P21* gene may result in alteration of P21 expression and/or activity, thereby modulating susceptibility to cancer (Keshava et al. 2002; Gravina et al. 2009; Ma et al. 2011).

In *MLH1* gene, a known exonic variant 37053568A > G (rs1799977) was observed in 10% of the study participants with the replacement of isoleucine to valine in codon 219 (219I > V) in exon 8 (Mathonnet et al. 2003). The polymorphism, I219V (A655G), was reported to be associated with childhood acute lymphoblastic leukemia (Listgarten et al. 2004). Studies also found a significant association between breast cancer and homozygous GG variant (Raptis et al. 2007). The homozygous or heterozygous G allele at nucleotide position 655 in *MLH1* gene was commonly reported for western populations (Christensen et al. 2008; Mann et al. 2008). However, in the current study, it was detected only in 10% of GC patients. In GC patients, the G allele frequency was 10%, higher than in controls which demonstrated a frequency of 0.5%, similar to data reported in Eastern Asians where the G allele frequency is reported to be approximately 2% (Trojan et al. 2002). The nucleotide position 655 is in conserved region thorough all mammals in exon 8. Earlier report published from the result of functional analyses that the homozygous or heterozygous G allele has efficient DNA repair activity (Raevaara et al. 2005; Kondo et al. 2003) and binding properties to PMS2 (Kim et al. 2004). It was well documented that 655A > G is also associated with reduced MLH1 protein expression in sporadic CRCs in Korean population (Marchetti et al. 1995). The HOPE analysis (in silico study) showed that the mutant type is smaller in size than the wild type, which might affect the intramolecular and external interactions due to clashes. 37088831C > G variant in intron 16 of *MLH1* gene was identified in 5% of GC samples. It is found to affect splice site by increasing the acceptor marginally (wt 0.5187/mu 0.5635) which might affect the normal splicing leading to abrupt transcription of this gene.

In the present study, *MDM2* gene had no significant mutations in GC samples. According to the previous studies, mutations and polymorphisms were identified in various exons of *MDM2* gene in esophageal and GC (Sauli et al. 2015). The known variant of *MDM2*, rs2279744 was found to influence independently the susceptibility to GC in Chinese population (Cho et al. 2008). According to Cho et al. (2008), SNPs of

MDM2 gene were not associated with increased GC risk in Korean population, and is consistent with our study. This can be explained due to the difference in the genetic pool and other demographic and dietary factors between the different populations.

The effect of mutations in the splice site was identified by Human Splicing Finder (HSF). We used all the intronic mutations and polymorphism that disturb the active site of 5′ss and 3′ss for validating the splicing effect and new cryptic splice site. The sequence of branch point represents another essential splicing signal. The analysis further revealed that 36653580C > T in 3′UTR of *P21*, branch point is changing with a CV variation of − 1.53% due to which the motif might get abolished leading to splicing defect (Table 5). 36653580C > T polymorphism is thought to cause a functional change in P21, and as this polymorphism lies in a crucial region for cell differentiation, and its proliferation may increase cancer risk by altering messenger RNA stability, which in turn may affect subsequent protein activity. Similarly, for *MLH1* intron 16, 37088831C > G mutation, the CV variation is − 4.94% conferring to splicing defect. Mutations located in the introns of mismatch repair genes can interfere with splicing and cause alternately spliced mRNA transcripts leading to non-functional mismatch repair proteins (Petersen et al. 2013).

According to the MDR analyses, the best model for GC risk in Mizo population is combination of smoked meat/vegetable and oral snuff consumption after performing the gene–environment interaction (Table 1, Fig. 2). Consumption of smoked meat showed the highest independent effect (20.56%) and also had modest synergistic interaction with sa-um (0.43%). Oral snuff (8.19%) also explained considerable entropy independently in GC risk and thus validated the results of gene-environment interaction. In India, many epidemiological studies reported significant positive association of tobacco and dietary habits containing harmful carcinogen such as *N*-nitroso compounds with GC (Sumathi et al. 2009). PAHs generated during preparation of heat-dried smoked food have been significantly associated in different geographical population in the world with high stomach cancer rates (Yeh et al. 2009). Soda-an alkaline preparation, frequently used as food additives, was significantly associated with increased risk of stomach cancer in Mizoram (Phukan et al. 2006). Different tobacco products such as oral snuff and smoking contains high amount of carcinogens like the nitrosamines which acts as cofactor for development of GC (IARC 2007).

We hypothesized that the balance in cell cycle control is disrupted by lifestyle habits (environmental risk factors) leading to the alteration (mutations) in the regulatory genes. These mutations affect the normal cell cycle by inducing abnormal distribution of DNA content that ultimately results in tumorigenesis. The aberrant content of DNA, or aneuploidy, is a hallmark of tumorigenesis (Giam and Rancati 2015). Thus, the flow cytometric study was conducted to evaluate the DNA content and S-phase fraction. Flow cytometric analysis of tumor samples showed 17.5% of aneuploid and 82.5% diploid for gastric tumor samples. A high S-phase fraction (SPF) was observed in GC samples (51.24–72.09) compared to controls (32.45–44.12) (Table 2, Fig. 4). The G2/M phase was significantly arrested in most of the GC tumor samples. In previous studies, DNA aneuploidy has been reported in 40–50% of GC tumors (Brito et al. 1993; Malumbres and Carnero 2003). Flow cytometric analysis predicted DNA aneuploid in 07 (17.5%) and diploid in 33 (82.5%) tumor samples (Table 4) followed by a high S-phase fraction (SPF). Deregulation of cell cycle events leads to uncontrolled cell proliferation and a high S-phase fraction which is a hallmark of GC (Baba et al. 2002). According to an earlier study, the *P21* variant genotypes have been demonstrated to play an important role in cell cycle control. The disruption in cell cycle control due to DNA damage is probably caused by carcinogens present in tobacco-related product (Flejou et al. 1993). Arrest of G2/M phase was observed in case of aneuploidy. DNA aneuploidy has been reported in 40–50% of tumors (Nanus et al. 1989; Quirke et al. 2005). Previous study reported aneuploidy in 76% (25 of 33) of adenocarcinomas arising in the gastric cardia, compared with 30% (8 of 27) of adenocarcinomas arising in the gastric antrum (Gleeson et al. 1998).

Conclusion

In conclusion, our findings indicate that mutations in cell cycle regulatory (*P21*) and mismatch repair (*MLH1*) genes are more predisposed to higher incidence of GC in Mizo population and may play an important role in tumorigenesis by inducing the aberrant distribution of DNA content during different phases of cell cycle in tumor cells. Ethnicity and dietary habits are acting as crucial covariates, suggesting that the mutations have different penetrance according to ethnicity, dietary, and lifestyle habits. This study could afford early detection of patients who are at risk of developing micro- or macroscopic, pathological lesions as well as the introduction of appropriate preventive measures. Due to the complexity as well as the correlations of multiple genetic and environmental factors in the development of GC, large population studies are required in order to overcome the limitation of sample size and encompass virtually all variables including the exposure to environmental factors, ethnic and demographic features besides the association with mutations in genes for DNA repair genes, cell cycle regulatory genes, and cell cycle study.

Funding This work was supported by DBT—Advanced Level State Biotech Hub (BT/04/NE/2009 dt. 29.08.2014), Bioinformatics Infrastructure Facility (No. BT/BI/12/060/2012 (BTISNeT), and DBT eLibrary Consortium (DeLCON) sponsored by the Department of

Biotechnology (DBT), New Delhi, Govt. of India, Mizoram University which provided all the essential facilities to carry out the work.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflicts of interest to report.

Ethics, consent, and permissions All participants gave written informed consent to the study protocol which was approved by the Ethical Committee of the Civil Hospital, Mizoram and Mizoram University, India (B.12018/1/13-CH(A)/IEC), to conduct and publish the research work. The study protocol was also approved by the Institutional Review Board of all institutes involved in the study.

References

An J-H, ASM J, Kim JW, Kim CH, Choi KH (2014) The expression of P21 is upregulated by forkhead box A1/2 in p53-null H1299 cells. FEBS Lett 588:4065–4070

Appelman HD, Mclaughlin JK, Blot WJ, Fraumeni JF (1992) A cohort study of stomach cancer in a high-risk American population. Cancer 69:2867–2868

Baba H, Korenaga D, Kakeji Y, Haraguchi M, Okamura T, Maehara Y (2002) DNA ploidy and its clinical implications in GC. Surgery 131(1 Suppl):S63–S70

Bahl R, Arora S, Nath N, Mathur M, Shukla NK, Ralhan R (2000) Novel polymorphism in P21(waf1/cip1) cyclin dependent kinase inhibitor gene: association with human esophageal cancer. Oncogene 19:323–328

Boland CR, Goel A (2010) Microsatellite instability in colorectal cancer. Gastroenterology 138(6):2073–2087

Bouska A, Lushnikova T, Plaza S, Eischen CM (2008) MDM2 promotes genetic instability and transformation independent of p53. Mol Cell Biol 28:4862–4874

Brito MJ, Filipe MI, Williams GT, Thompson H, Ormerod MG, Titley J (1993) DNA ploidy in early gastric carcinoma (T1): a flow cytometric study of 100 European cases. Gut 34:230–234

Campbell PT, Curtin K, Ulrich CM, Samowitz WS, Bigler J, Velicer CM, Caan B, Potter JD, Slattery ML (2009) Mismatch repair polymorphisms and risk of colon cancer, tumour microsatellite instability and interactions with lifestyle factors. Gut 58:661–667

Campos F, Carrasquilla G, Koriyama C, Serra M, Carrascal E, Itoh T, Nomoto M, Akiba S (2006) Risk factors of GC specific for tumor location and histology in Cali, Colombia. World J Gastroenterol 12:5772–5779

Cattaert T, Calle ML, Dudek SM, Mahachie JJM, Van-Lishout F et al (2011) Model-based multifactor dimensionality reduction for detecting epistasis in case–control data in the presence of noise. Ann Hum Genet 75(1):78–89

Chao A, Thun MJ, Henley SJ, Jacobs EJ, Mccullough ML, Calle EE (2002) Cigarette smoking, use of other tobacco products and stomach cancer mortality in us adults: the cancer prevention study II. Int J Cancer 101:380–389

Chen C, Pearson AM, Gray JI (1990) Meat mutagens. Adv Food Nutr Res 34:387–449

Cho YG, Choi BJ, Song JH, Kim CJ, Cao Z, Nam SW, Lee JY, Park WS (2008) No association of MDM2 T309G polymorphism with susceptibility to Korean GC patients. Neoplasma 55:256–260

Choudhury JH, Singh SA, Kundu S, Choudhury B, Talukdar FR, Srivasta S et al (2015) Tobacco carcinogen-metabolizing genes CYP1A1, GSTM1, and GSTT1 polymorphisms and their interaction with tobacco exposure influence the risk of head and neck cancer in northeast Indian population. TumorBiol 36(8):5773–5783

Christensen LL, Madsen BE, Wikman FP, Wiuf C, Koed K, Tjonneland A (2008) The association between genetic variants in hMLH1 and hMSH2 and the development of sporadic colorectal cancer in the Danish population. BMC Med Genet 9:52–63

Ciccarelli C, Marampon F, Scoglio A, Mauro A, Giacinti C, Cesaris PD (2005) P21WAF1 expression induced by MEK/ERK pathway activation or inhibition correlates with growth arrest, myogenic differentiation and onco-phenotype reversal in rhabdomyosarcoma cells. Mol Cancer 13(4):14

Correa P, Fontham E, Pickle LW, Chen V, Lin YP, Haenszel W (1985) Dietary determinants of GC in south Louisiana inhabitants. J Natl Cancer Inst 75:645–654

Curado MP, Edwards B, Shin HR, Storm H, Ferlay J, Heanue M, et al. (eds) (2007) Cancer incidence in five continents, vol. IX, IARC Scientific Publications No. 160. IARC, Lyon

Decesse JT, Medjkane S, Datto MB, Crémisi CE (2001) RB regulates transcription of the P21/WAF1/CIP1 gene. Oncogene 20(8):962–971

Dikshit R, Gupta PC, Ramasundarahettige C, Gajalakshmi V, Aleksandrowicz L, Badwe R (2012) Million death study collaborators. Cancer mortality in India: a nationally representative survey. Lancet 379:1807–1816

Ferlay J, Soerjomataram I, Dikshit R, Eser S, Mathers C, Rebelo M (2015) Cancer incidence and mortality worldwide, sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer 136:359–386

Flejou JF, Potet F, Muzeau F, Le-Pelletier F, Fekete F, Henin D (1993) Overexpression of p53 protein in Barrett's syndrome with malignant transformation. J Clin Pathol 46:330–333

Furberg H, Lichtenstein P, Pedersen NL, Bulik C, Sullivan PF (2006) Cigarettes and oral snuff use in Sweden: prevalence and transitions. Addiction 101(10):1509–1515

Gartel AL (2005) The conflicting roles of the CDK inhibitor P21(CIP1/WAF1) in apoptosis. Leuk Res 29:1237–1238

Gartel AL, Tyner AL (1999) Transcriptional regulation of the P21 (WAF/CIP1) gene. Exp Cell Res 246:280–289

Ghatak S, Muthukumaran RB, Nachimuthu SK (2013) A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech 24:224–231

Ghatak S, Yadav RP, Lalrohlui F, Chakraborty P, Ghosh S, Ghosh S (2016) Xenobiotic pathway gene polymorphisms associated with GC in high risk Mizo-mongoloid population, northeast India. Helicobacter 21:523–535

Ghatak S, Zothansanga PJL, Nachimuthu SK (2014) Co-extraction and PCR based analysis of nucleic acids from formalin-fixed paraffin-embedded specimens. J Clin Lab Anal 29(6):485–492

Giam M, Rancati G (2015) Aneuploidy and chromosomal instability in cancer: a jackpot to chaos. Cell Div 10:3

Gleeson CM, Sloan JM, McManus DT, Maxwell P, Arthur K, McGuigan JA (1998) Comparison of p53 and DNA content abnormalities in adenocarcinoma of the oesophagus and gastric cardia. Br J Cancer 77(2):277–286

Gooding C, Clark F, Wollerton MC, Grellscheid SN, Groom H, Smith CW (2006) A class of human exons with predicted distant branch points revealed by analysis of AG dinucleotide exclusion zones. Genome Biol 7:23–31

Gravina S, Lescai F, Hurteau G, Brock GJ, Saramaki A, Salvioli S (2009) Identification of single nucleotide polymorphisms in the P21 (CDKN1A) gene and correlations with longevity in the Italian population. Aging (Albany NY) 1:470–480

Guerrette S, Acharya S, Fishel R (1999) The interaction of the human MutL homologues in hereditary nonpolyposis colon cancer. J Biol Chem 274:6336–6341

Hahn LW, Ritchie MD, Moore JH (2003) Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. Bioinformatics 19(3):376–382

Hecht SS (2003) Tobacco carcinogens, their biomarkers and tobacco-induced cancer. Nat Rev Cancer 3:733–744

Hedley DW (1994) DNA analysis from paraffin-embedded blocks. Flow cytometry Second Edition, Part A PP: 231–240

Igaki H, Sasaki H, Kishi T, Sakamoto H, Tachimori Y, Kato H (1994) Highly frequent homozygous deletion of the p16 gene in esophageal cancer cell lines. Biochem Biophys Res Commun 203:1090–1095

Ihsan R, Devi TR, Yadav DS, Mishra AK, Sharma J, Zomawia E (2011) Investigation on the role of p53 codon 72 polymorphism and interactions with tobacco, betel quid, and alcohol in susceptibility to cancers in a high-risk population from north east India. DNA Cell Biol 30:163–171

Jung YS, Qian Y, Chen X (2010) Examination of the expanding pathways for the regulation of P21 expression and activity. Cell Signal 22(7): 1003–1012

Keshava C, Frye BL, Wolff MS, McCanlies EC, Weston A (2002) Waf-1 (P21) and p53 polymorphisms in breast cancer. Cancer Epidemiol Biomark Prev 11:127–130

Kim JC, Roh SA, Koo KH, Ka IH, Kim HC, Yu CS (2004) Genotyping possible polymorphic variants of human mismatch repair genes in healthy Korean individuals and sporadic colorectal cancer patients. Fam Cancer 3:129–137

Kneller RW, Guo WD, Hsing AW, Chen JS, Blot WJ, Li JY et al (1992) Risk factors for stomach cancer in sixty-five Chinese counties. Cancer Epidemiol Biomark Prev 1:113–118

Kondo E, Horii A, Fukushige S (2001) The interaction domains of three MutL heterodimers in man: hMLH1 interacts with 36 homologous amino acid residues within hMLH3, hPMS1 and hPMS2. Nucleic Acid Res 29:1695–1708

Kondo E, Suzuki H, Horii A, Fukushige S (2003) A yeast two-hybrid assay provides a simple way to evaluate the vast majority of hMLH1 germ-line mutations. Cancer Res 63:3302–3308

Lin G, Fang F, Yu XJ, Yu L (2011) Meta-analysis of the relationship between P21 Ser31Arg polymorphism and lung cancer susceptibility. Genet Mol Res 10:2449–2456

Lipkin SM, Wang V, Jacoby R, Banerjee-Basu S, Baxevanis AD, Lynch HT (2000) MLH3: a DNA mismatch repair gene associated with mammalian microsatellite instability. Nat Genet 24:27–35

Listgarten J, Damaraju S, Poulin B, Cook L, Dufour J, Driga A (2004) Predictive models for breast cancer susceptibility from multiple single nucleotide polymorphisms. Clin Cancer Res 10:2725–2737

Ma H, Zhou Z, Wei S, Wei Q (2011) Association between P21 Ser31Arg polymorphism and cancer risk: a meta-analysis. Chin J Cancer 30: 254–263

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

Malakar M, Devi KR, Phukan RK, Kaur T, Deka M, Lalhriat-Puia L (2012) Genetic polymorphism of glutathione S-transferases M1 and T1, tobacco habits and risk of stomach cancer in Mizoram, India. Asian Pac J Cancer Prev 13(9):4725–4732

Malumbres M, Carnero A (2003) Cell cycle deregulation: a common motif in cancer. Prog Cell Cycle Res 5:5–18

Mann A, Hogdall E, Ramus SJ, DiCioccio RA, Hogdall C, Quaye L (2008) Mismatch repair gene polymorphisms and survival in invasive ovarian cancer patients. Eur J Cancer 44:2259–2265

Manuguerra M, Matullo G, Veglia F, Autrup H, Dunning AM, Garte S (2007) Multi-factor dimensionality reduction applied to a large prospective investigation on gene-gene and gene-environment interactions. Carcinogenesis 28(2):414–422

Marchetti A, Buttitta F, Pellegrini S, Merlo G, Chella A, Angeletti C (1995) MDM2 gene amplification and overexpression in non-small cell lung carcinomas with accumulation of the p53 protein in the absence of p53 gene mutations. Diagn Mol Pathol 4:93–97

Mathonnet G, Krajinovic M, Labuda D, Sinnett D (2003) Role of DNA mismatch repair genetic polymorphisms in the risk of childhood acute lymphoblastic leukaemia. Br J Haematol 123:45–48

Merkel DE, Mcguire WL (1990) Ploidy, proliferative activity and prognosis. Cancer 65:1194–1205

Mousses S, Ozcelik H, Lee PD, Malkin D, Bull SB, Andrulis IL (1995) Two variants of the CIP1/WAF1 gene occur together and are associated with human cancer. Hum Mol Genet 4:1089–1092

Nanus DM, Kelsen DP, Niedzwiecki D, Chapman D, Brennan M, Cheng E (1989) Flow cytometry as a predictive indicator in patients with operable GC. J Clin Oncol 7(8):1105–1112

National Cancer Registry Programme (2013) Three-year report of the population based cancer registries 2011–2013. National cancer registry programme, Indian Council of Medical Research (ICMR), Bangalore, India. Available from, http://www.pbcrin-dia.org. Accessed 16 Oct 2014

Nomura A, Grove JS, Stemmermann GN, Severson RK et al (1990) A prospective study of stomach cancer and its relation to diet, cigarettes, and alcohol consumption. Cancer Res 50:627–631

Oren M, Damalas A, Gottlieb T, Michael D, Taplick J, Leal JF (2002) Regulation of p53: intricate loops and delicate balances. Biochem Pharmacol 64:865–871

Parker SB, Eichele G, Zhang P, Rawls A, Sands AT, Bradley A (1995) p53-independent expression of P21Cip1 in muscle and other terminally differentiating cells. Science 267(5200):1024–1027

Petersen SM, Dandanell M, Rasmussen LJ, Gerdes AM, Krogh LN, Bernstein I (2013) Functional examination of MLH1, MSH2, and MSH6 intronic mutations identified in Danish colorectal cancer patients. BMC Med Genet 14:103–114

Phillips DH (1999) Polycyclic aromatic hydrocarbons in the diet. Mutat Res 443:139–147

Phukan RK, Hazarika NC, Baruah D, Mahanta J (2004) High prevalence of stomach cancer among the people of Mizoram, India. Curr Sci 87: 285–286

Phukan RK, Narain K, Zomawia E, Hazarika NC, Mahanta J (2006) Dietary habits and stomach cancer in Mizoram, India. J Gastroenterol 41:418–424

Phukan RK, Zomawia E, Narain K, Hazarika NC, Mahanta J (2005) Tobacco use and stomach cancer in Mizoram, India. Cancer Epidemiol Biomark Prev 14:1892–1896

Piazuelo MB, Correa P (2013) Gastric cáncer: overview. Colomb Med (Cali) 44(3):192–201

Quirke DP, Dixon MF, Clayden AD, Durdey P, Dyson JED, Williams NS (2005) Prognostic significance of DNA aneuploidy and cell proliferation in rectal adenocarcinomas. J Pathol 151(4):285–291

Raevaara TE, Korhonen MK, Lohi H, Hampel H, Lynch E, Lönnqvist KE (2005) Functional significance and clinical phenotype of nontruncating mismatch repair variants of MLH1. Gastroenterology 129:537–549

Raptis S, Mrkonjic M, Green RC, Pethe VV, Monga N, Chan YM (2007) MLH1 -93G>A promoter polymorphism and the risk of microsatellite-unstable colorectal cancer. J Natl Cancer Inst 99:463–474

Ritchie MD, Hahn LW, Moore JH (2003) Power of multifactor dimensionality reduction for detecting gene-gene interactions in the presence of genotyping error, missing data, phenocopy, and genetic heterogeneity. Genet Epidemiol 24(2):150–157

Santarelli RL, Pierre F, Corpet DE (2008) Processed meat and colorectal cancer: a review of epidemiologic and experimental evidence. Nutr Cancer 60(2):131–144

Sauli E, Hongna L, Vedastus AK, Weiyue H, Song L, Nongyue H (2015) Polymorphisms in NEIL-2, APE-1, CYP2E1 and MDM2 genes are independent predictors of GC risk in a northern Jiangsu population (China). J Nanosci Nanotechnol 15(7):4815–4828

Shiohara M, Deiry WE, Wada M, Nakamaki T, Takeuchi S, Yang R (1994) Absence of WAF1 mutations in a variety of human malignancies. Nature 84:3781–3784

Stepanov I, Hecht SS, Ramakrishnan S, Gupta PC (2005) Tobacco-specific nitrosamines in smokeless tobacco products marketed in India. Int J Cancer 116(1):16–19

Stepanov I, Villalta PW, Knezevich A, Jensen J, Hatsukami DK, Hecht SS (2010) Analysis of 23 polycyclic aromatic hydrocarbons in smokeless tobacco by gas chromatography–mass spectrometry. Chem Res Toxicol 23(1):66–73

Sumathi B, Ramalingam S, Navaneethan U, Jayanthi V (2009) Risk factors for GC in South India. Singap Med J 50:147–151

Torre L, Bray F, Siegel RL, Ferlay J, Lortet-Tieulent J, Jemal A (2015) Global Cancer Statistics, 2012. CA Cancer J Clin 65:87–108

Trojan J, Zeuzem S, Randolph A, Hemmerle C, Brieger A, Raedle J (2002) Functional analysis of hMLH1 variants and HNPCC-related mutations using a human expression system. Gastroenterology 122:211–219

Tsugane S, Sasazuki S (2007) Diet and the risk of GC: review of epidemiological evidence. Gastric Cancer 10:75–83

Tsugane S, Sasazuki S, Kobayashi M, Sasaki S (2004) Salt and salted food intake and subsequent risk of GC among middle-aged Japanese men and women. Br J Cancer 90:128–134

Wang X, Wang J, Huang V, Place RF, Li LC (2012) Induction of NANOG expression by targeting promoter sequence with small activating RNA antagonizes retinoic acid-induced differentiation. Biochem J 443:821–828

Ward MH, Lopez-Carrillo L (1999) Dietary factors and the risk of GC in Mexico City. Am J Epidemiol 149:925–392

Watanabe H, Fukuchi K, Takagi Y, Tomoyasu S, Tsuruoka N, Gomia K (1995) Molecular analysis of the Cip1/Waf1 (P21) gene in diverse types of human tumors. Biochim Biophys Acta 1263:275–280

Yaghoobi M, Bijarchi R, Narod SA (2010) Family history and the risk of GC. Br J Cancer 102(2):237–242

Yan L, Fang L, Sha Z, Suqing S, Li L, Lifeng L (2015) Genetics and GC susceptibility. Int J Clin Exp Med 8(6):8377–8383

Yeh JM, Kuntz KM, Ezzati M, Goldie SJ (2009) Exploring the cost-effectiveness of *Helicobacter pylori* screening to prevent GC in China in anticipation of clinical trial results. Int J Cancer 124:157–166

Mitochondrial complex I and V gene polymorphisms associated with breast cancer in mizo-mongloid population

Article *in* Breast Cancer · April 2015

DOI: 10.1007/s12282-015-0611-1 · Source: PubMed

CITATIONS
5

READS
103

7 authors, including:

Souvik Ghatak
Mizoram University
30 PUBLICATIONS **182** CITATIONS

SEE PROFILE

Nachimuthu Dr
Periyar University
85 PUBLICATIONS **296** CITATIONS

SEE PROFILE

Breast Cancer
DOI 10.1007/s12282-015-0611-1

ORIGINAL ARTICLE

Mitochondrial complex I and V gene polymorphisms associated with breast cancer in mizo-mongloid population

Sunaina Thapa[1] · Freda Lalrohlui[1] · Souvik Ghatak[1] · John Zohmingthanga[2] · Doris Lallawmzuali[3] · Jeremy L. Pautu[3] · Nachimuthu Senthil Kumar[1]

Received: 29 January 2015 / Accepted: 7 April 2015

Abstract
Background Mizoram has the highest incidence of cancer in India. Among women, breast cancer is most prevalent and the state occupies fifth position globally. The reason for high rate of cancer in this region is still not known but it may be related to ethnic/racial variations or lifestyle factors.
Methods The present study aims to identify the candidate mitochondrial DNA (mtDNA) biomarkers—ND1and ATPase for early breast cancer diagnosis in Mizo population. Genomic DNA was extracted from blood samples of 30 unrelated breast cancer and ten healthy women. The mtNDI and mtATP coding regions were amplified by step-down PCR and were subjected to restriction enzyme digestion and direct sequencing by Sanger method. Subsequently, the results of the DNA sequence analysis were compared with that of the revised Cambridge Reference Sequence (rCRS) using Mutation Surveyor and MITOMAP.
Results Most of the mutations were reported and new mutations that are not reported in relationship with breast cancer were also found. The mutations are mostly base substitutions. The effect of non-synonymous substitutions on the amino acid sequence was determined using the PolyPhen-2 software. Statistical analysis was performed for both cases and controls. Odds ratios (ORs) and 95 % confidence intervals (CIs) were estimated from logistic regression. High intake of animal fat and age at menarche was found to be associated with a higher risk of breast cancer in Mizo population.
Conclusion Our results also showed that ATPase6 as compared to ATPase8 gene is far more predisposed to variations in Mizo population with breast cancer and this finding may play an important role in breast cancer prognosis.

Keywords mtDNA · ATP6 · ND1 · ATP8 · OXPHOS · Breast cancer

Introduction

Mizoram is the southernmost state in the northeast, sharing its borders with Tripura, Assam, and Manipur along with Myanmar and Bangladesh. The Mizo people are believed to be a part of the Mongoloid race and Mizoram has the highest number of tribal people among all states of India. According to the National Cancer Registry Program 2010, Mizoram has the highest incidence of cancer in India. Among women, breast and cervix cancer are most prevalent and the state occupies fifth position globally [1].

Breast cancer is a type of cancer which originates from breast tissues and is a heterogeneous disease with a variety of subtypes, with distinct gene expression profiles, that have substantial implications for prognosis and survival rates [2]. Most of the breast cancers are like some solid tumors, approximately 25–30 % are found in familial

Electronic supplementary material The online version of this article (doi:10.1007/s12282-015-0611-1) contains supplementary material, which is available to authorized users.

✉ Nachimuthu Senthil Kumar
nskmzu@gmail.com

[1] Department of Biotechnology, Mizoram University, Aizawl 796004, Mizoram, India

[2] Department of Pathology, Civil Hospital, Aizawl 796001, Mizoram, India

[3] Mizoram State Cancer Institute, Zemabawk, Aizawl 796017, Mizoram, India

Published online: 21 April 2015

clustering and 5–10 % are found due to a significant hereditary component [3]. Yeole et al. [4] described breast cancer as an alarming health problem in India and according to many reports breast cancer has been a major concern for the women population in India. According to National Cancer Registry Program (NCRP), a recent report for 2008 found that in North Eastern states alone, the breast cancer incidence was reported to be 23.6 % of the total cancers [5]. Northeast India has always been one of the hotspot regions for breast cancer because of certain factors such as genotoxic stress from tobacco exposure [6]. Several environmental or dietary risk factors that may contribute to or hasten the development of breast cancer have been identified, including mainly lifestyle and reproductive factors. The strongest risk factor for breast cancer is a family history of breast and/or ovarian cancer, the associated risk being even higher for family history of early-onset disease (≤age 40) [7].

Mitochondrial genes are intensively studied nowadays and it has been proven to be a powerful and efficient molecular genetics tool for investigating the (maternal) genetic history of human populations [8]. Mitochondria are organelles of high importance for the generation of ATP through the oxidative phosphorylation. As compared to nDNA, human mtDNA is exclusively small (16569 bp) which encodes 13 proteins that comprise four enzyme complexes of the respiratory chain (Complex I, III, IV, and ATP synthase i.e., Complex V), and genes specifying 22 transfer RNAs, and 2 ribosomal RNAs (12S and 16S) that are components of mitochondrial protein synthesizing system [9]. Several reports have found an association between somatic mtDNA mutations and the development, progression, or metastasis of cancer and one of the contributing factors that lead to the development of cancer was found to be inherited mtDNA polymorphisms [10–12]. Many earlier reports had found mtDNA mutations in different types of cancer, such as colon, breast, pancreatic, prostate, and other several cancers [13]. Studies on mitochondrial ATPase and ND genes in the field of oxidative phosphorylation have converged in the last few years. The mtND1 (NADH dehydrogenase subunit 1) gene which encompasses 955 base pairs coding sequences encodes one of the seven subunits of respiratory complex 1 and takes part in the first initial step of the electron transport chain of the mitochondrial energy-generating pathway, OXPHOS whereas ATPase (Complex V) genes are essential for ATP production and play a vital role in the apoptosis pathways [14].

Therefore, this study was undertaken to evaluate mitochondrial ND1 and ATPase alterations and to study the lifestyle factors that are associated with breast cancer patients in Mizo population.

Materials and methods

Sample collection

Blood samples were collected from breast cancer patients ($n = 30$) randomly and healthy controls ($n = 10$; age and sex-matched) from Mizoram State Cancer Institute, Aizawl. All the patients involved in this study were medically confirmed primary breast cancer patients who had given written consent to participate in this study. All the samples belong to invasive ductal carcinoma breast cancer subtype. Peripheral blood samples (5 ml) were collected using venipuncture procedure and stored in EDTA vacutainers from all the subjects and stored at −80 °C for further use. Also information on demographic factors, menstrual and reproductive history, hormone use, dietary habits, previous disease history, physical activity, tobacco and alcohol use, weight, and family history of cancer was collected during an in-person interview by trained study interviewers using a structured questionnaire (Table 1). The protocol of our study was approved by the Institutional Review Board of all institutes involved in the study. All the volunteers were informed about the study beforehand and their full consent was taken before their participation.

DNA isolation form blood samples

Genomic DNA was extracted from peripheral blood samples of breast cancer patients by lysing the RBCs using a Lysis buffer such as ammonium bicarbonate and ammonium chloride, which is a hypotonic buffer, having an eligible lysing effect on lymphocytes. Thrice the volume of RBC lysis buffer was added to the blood samples, mixed thoroughly by vortexing and inverting for 5 min and centrifugation for 10 min at 5000 rpm. The supernatant was discarded and to the pellet, again three volumes of RBC lysis buffer was added and vortexed as above. The sample was mixed thoroughly by inverting and centrifuged again at 5000 rpm repeatedly at least for 2–3 times till the supernatant and a white pellet were obtained. The supernatant was decanted completely and the pellet was resuspended in 700 µl PBS, vortexed and centrifuged again for 10 min at 5000 rpm, followed by the addition of 500 µl amount of cell lysis buffer (100 mM Tris–HCl, 50 mM EDTA, 50 mM Nacl 10 % SDS, pH 7.5) and 10 µl of Proteinase K (10 mg/ml stock) (Hi-media). The samples were vortexed to completely dissolve the pellet and the samples were incubated for lysis in a water bath at 56 °C for 2 h. The tubes were shaken and an equal volume of PCI was added and mixed thoroughly for 1 min by inverting. Centrifugation at 10000 rpm for 10 min at 4 °C was done and to a new tube the aqueous upper layer was transferred which

Table 1 Demographic and reproductive information of the patients covered in this study

Status	Age (years)	Weight (Kg)	Diet information			Habits					Reproductive information				Medical History	Familial History of Breast cancer	Familial History of other cancers
			Meat	Fat	Salt	Cigarette	KUHVA	Sahdah	Alcohol	Tuibur	Breast feeding	No. of child	Age of menarche (years)	Age of menopause (years)			
Cancer	24	50	+	++	+	−	+	+	−	−	+	1	14	–	+	−	+
	43	56	++	+++	++	+	+	+++	−	++	+	4	15	–	+	−	−
	37	50	+	+	+	−	−	−	+	−	+	5	18	–	+	+	−
	36	51	+	+	+	−	++	−	+	++	−	0	12	–	+	−	+
	32	55	++	+	+	−	−	−	−	−	+	4	15	–	−	−	+
	33	52	+++	+	+	++	+	−	++	−	+	1	14	–	+	−	+
	51	53	++	++	++	++	++	−	−	−	+	3	14	46	−	−	+
	75	43	+	++	++	+	−	−	−	++	+	2	17	50	+	−	−
	53	50	++	+	+	−	−	++	−	+	+	2	14	49	+	−	+
	73	52	+	+	+	−	−	−	−	−	+	6	15	50	−	−	−
	60	55	++	++	++	−	−	++	−	−	−	0	14	50	+	−	+
	40	46	+	+	+	++	++	++	−	−	+	6	16	−	−	−	+
	60	58	+++	+	+	−	−	−	−	−	+	2	16	53	−	−	−
	38	58	++	+	+	−	−	−	−	−	+	0	14	−	−	−	+
	62	56	++	+	+	++	+	+	−	+	+	5	16	50	−	−	−
	43	63	++	++	++	−	−	−	−	−	+	4	12	50	+	−	−
	52	47	++	++	++	−	+	−	+	+	+	3	15	−	+	−	+
	43	52	++	+++	+++	+	++	++	+	+	+	3	16	−	−	−	+
	45	52	++	++	++	++	++	+	−	−	+	4	13	45	−	−	+
	40	54	+	+	+	+	+	++	−	−	+	2	14	−	+	−	−
	78	53	+	+++	+++	+	−	−	−	−	+	3	14	54	+	−	−
	57	62	++	+	+	−	−	−	−	−	−	7	15	51	−	−	−
Healthy	50	52	+	+	+	+	+	+	−	−	+	1	16	48	+	−	+
	52	48	+	+	++	−	+	−	−	−	+	2	15	46	−	−	−
	48	52	++	+	+	+	++	+	−	−	+	2	14	47	−	−	−
	56	56	++	+	+	+	+	++	−	+	+	3	14	50	+	−	−
	48	52	+	+	++	−	−	+	−	−	+	2	15	48	+	+	+
	50	57	+	+	+	+	+	−	−	−	+	2	16	49	−	+	−
	54	51	+	+	+	−	−	−	−	−	+	2	16	48	+	−	−
	54	53	+	+	+	−	+	+	−	+	+	1	14	47	+	−	−

+ low; ++ medium; +++ High consumption; Familial Information: + Present; − Absent; Breast feeding: + Yes; − No; Age of menopause:–Not yet menopause

contained equal volumes (1:1) of phenol and chloroform: isoamyl alcohol in the ratio of 24:1. The contents were mixed thoroughly by inverting for 1 min and centrifuged again at 10000 rpm for 10 min at 4 °C.

The supernatant was transferred to a fresh microcentrifuge tube and two times the volume of isopropanol (Merck) was added and inverted gently for 10 times. Sodium acetate (50 µl) was added and incubated at −20 °C for 1 h then centrifuged at 13000 rpm for 10 min at 4 °C. After decanting the supernatant, 250 µl of 70 % ethanol was added and the pellet was gently tapped, followed by centrifugation at 13000 rpm for 10 min, and gently decanting the supernatant. The pellet was air dried in a laminar air flow followed by re-suspension in 50 µl of nuclease-free water or 1X TE (10 mM Tris–HCL, 1 mM EDTA, pH 7.6) buffer and stored in −20 °C [15]. Extracted DNA was tested using 0.8 % agarose gel electrophoresis. Yields and purity of DNA samples were estimated using spectrophotometer.

PCR amplification of the mtDNA ND1 and ATPase gene

The entire mitochondrial ND1 gene spanning across nucleotide positions 3307–4263 was amplified. The following primers were used for PCR amplifications, forward primer 5′-GAGCCCGGTAATCGCATAA-3′ and reverse primer 5′-GATAGGTGGCACGGAGAAT-3′. PCR reactions were performed in a total reaction volume of 25 µL containing 9.5 µL distilled water, 12.5 µL (1X) GoTaq® Green Master Mix (buffer pH 8.5), Taq polymerase, 200 mM each of dATP, dCTP, dGTP, dTTP, and 1.5 mM of $MgCl_2$ provided by (Promega, USA), 0.5 µL (10 pmol/µL) of each primer (Forward and Reverse), and 2 µL of 50 ng template DNA. The mixture was incubated in (Multigene-TM Gradient Thermal Cycler, Labnet International, Korea) with the following cycling conditions: an initial denaturation at 95 °C for 5 min, followed by 30 cycles of denaturation at 95 °C for 60 s, annealing of primers at 58 °C for 40 s, extension at 72 °C for 70 s, and a final extension cycle 72 °C for 5 min. A negative control reaction was prepared with all amplifications to ensure the reliability of results. PCR products were electrophoresed in 2 % agarose along with 100 bp DNA ladder.

The mtDNA ATPase region was amplified by PCR using primers KatPase-F (5′-CTAGAGCCCACTGTAA AGCTAAC-3′) and KatPase-R (5′GAGCGTTATGG AGTGGAAGT-3′). The ATPase polymerase chain reaction (PCR) was carried out in 25 µl total reaction volumes, each containing 100 ng of template DNA, 0.2 pmol of each primer, 2.5 µl of PCR buffer with 10X concentration, 1.5 mM $MgCl_2$, 200 mM dNTPs, and 1 unit of Taq DNA polymerase (Fermentas, Germany). The reaction mixture was heated to 95 °C for 5 min, which was followed by other 30 cycles each consisting of 1 min denaturation at 95 °C, 40 s annealing at 56 °C, 1 min of extension at 72 °C, and a final 5 min extension at 72 °C. The PCR amplification products (10 µl) were subjected to electrophoresis in a 1 % agarose gel in 1X TAE buffer at 80 V for 30 min, stained with (0.2 µg/ml) Ethidium Bromide and images were obtained in gel documentation system.

RFLP Screening

PCR amplification products were subjected to digestion reaction with HaeIII, AciI, TaqI, and RsaI enzymes. Ten µl of digestion reaction consists of: {5.6 µl of sterile distilled water, 1 µl of restriction buffer (10X), 0.4 µl of restriction enzyme} and 3 µl of amplified DNA. Then the digestion reaction was incubated at 37 °C for HaeIII, AciI, and RsaI and 56 °C for 3 h, 8–10 h, 6–10 h, and 10–12 h, respectively. Digestion products were subjected to electrophoresis in 8 % polyacrylamide gel in a 10X TBE buffer at a voltage of 40 V for 30 min later changed to 60 V and stained with 1 µl of ethidium bromide. The resulting fragments were visualized under the UV fluorescence. Gel images were captured using a gel documentation system. PCR products were purified and stored at −20 °C until sequencing at SciGenom Lab, Cochin, India.

Sequencing and Sequence analysis

All PCR products were sequenced from the opposite direction to ensure reading accuracy. Analysis of the sequence was performed using FinchTV 1.4 software (Geospiza, Inc., USA) and DNA Baser Sequence Assembler v 3.2 (2012) software. The sequences were compared with the latest version of Revised Cambridge Reference Sequence (rCRS) of the human mitochondrial DNA using the sequence analysis tool BLAST (http://www.ncbi.nlm.nih.gov/blast) (NCBI, Bethesda, USA). Mitochondrial genome sequence variations were identified using the Mitomap database. The results of the DNA sequence analysis were compared with the published Cambridge Sequence using Mutation Surveyor version 1.4 DNA mutation analysis software (Softgenetics, State College, PA). Sequence differences between cancer and healthy blood samples were recorded as mtDNA polymorphisms. Each polymorphism was then verified against the Mitomap database (http://www.mitomap.org/). Sequence variants that were not found in that database were classified as new polymorphism/mutation, whereas others already reported were classified as reported polymorphism/mutation. Sequences obtained from the present study were submitted to the EMBL-EBI database and the accession numbers are shown in Table 2.

Table 2 EBI Accession numbers of samples sequenced for mutation analysis

Gene name	EBI accession number	Sample type
ATP 8	LN558468	Breast cancer
	LN558469	
	LN558470	
	LN558471	
	LN558472	
	LN558473	
	LN558474	
	LN558448	Healthy
	LN558449	
	LN558450	
	LN558451	
	LN558452	
ATP6	LN558479	Breast cancer
	LN558480	
	LN558481	
	LN558482	
	LN558483	
	LN558484	
	LN558485	
	LN558464	Healthy
	LN558465	
	LN558466	
	LN558467	
NDI	LN558490	Breast cancer
	LN558491	
	LN558492	
	LN558493	
	LN558494	
	LN558495	
	LN558496	
	LN558510	Healthy
	LN558511	
	LN558512	

Pathogenicity prediction by protein amino acid change

Non-synonymous (coding) substitutions and its impact in the resulting protein was measured using PolyPhen-2 (v. 2.2.2) software, which predicts the probable impact of an amino acid substitution variation on the structure and function of the corresponding protein, which is inferred as benign and damaging effects.

Statistical Analysis

Case–control study was done for Hardy–Weinberg equilibrium by a Chi-square (χ^2) test with one degree of freedom (*df*). The demographic factors and habits between cases and controls were compared using the χ^2 test with Yates' correction or Fisher's exact test. Odds ratios with 95 percent confidence intervals were calculated for both cases and controls using logistic regression models that examined the association between breast cancer status and risk factors such as age of menarche and menopause, breast feeding, fat intake, smoking, alcohol consumption, familial cases of breast cancer, and other cancers. An RR was considered statistically significant if the 95 % CI did not contain 1.00. For all tests, a two-sided P value < 0.05 was considered statistically significant. All statistical analyses were performed using SPSS 20.0 for Windows (SPSS Ibérica, Madrid and Spain) and SigmaPlot 12.00, SYSTAT 13.0. (Systat Software Inc., USA).

Genetic analysis

The breast cancer and normal sequences were carefully checked and aligned using the Clustal W algorithm. The number of polymorphic sites, parsimony informative sites, the rate of transitions/transversions, and the nucleotide frequency were calculated using MEGA 6. The genetic distance between healthy and breast cancer was measured with MEGA 6 using the Kimura 2-parameter model. For the HKY model, we first created maximum estimates of parameter k (transition/transversion rate ratio) and a (gamma-shaped parameter) and likelihood value using the SA algorithm with the Jukes and Cantor's (JC) model and obtained ML nucleotide frequencies. Codon positions included were 1st + 2nd + 3rd + Noncoding. All positions containing gaps and missing data were eliminated. The variable substitution site was calculated by the formula:

$$H_i = \left(\sum_{i=1}^{4} Pj \log_2 Pj \right).$$

The variability at site '*i*' is measured by the entropy (or information), where '*j*' = 1, 2, 3, 4 corresponding to nucleotide A, C, G, and T, and 'pj' is the proportion of nucleotide '*j*' at site '*i*'. If all nucleotides at site '*i*' are identical, then 'Hi' = 0. Substitutions will lead to polymorphic sites at which the 'H' value will be larger than 0. It is 'Hi' that is plotted over sites.

Results

Demographic characteristics and some reproductive factors of breast cancer patients are shown in Table 1. From the demographic results, the cancer patients under study were of different age groups (ranging from 24 to 78 years), were mostly married and had children. Most of them breast-fed

up to more than 6 months and they had varying age of menarche and menopause. However, they mostly had early menarche, below 16 years. From our study, it was found that the habitual diet is high in animal fats in the form of saum (fermented pork fat), red and smoked meats whereas low intake of fruits among this group (data not shown). And high intake of animal fat is found to be one of the most major risk factor for breast cancer in our study (Relative Risk, RR: 2.08; Odd Ratio, OR: 1.32, 95 % CI 0267–4.02). The increased risk was correlated with high intake of animal fats, in the form of pork meat and saum (fermented pork fat). Age at menopause is also found as one of the significant risk factors in this population (RR = 1.56; Odd Ratio, OR: 1.06, 95 % CI 4.067–11.73). A slight increased risk of breast cancer was also observed in women with other types of cancers in their family (RR = 0.93; OD: 1.07; 95 % CI = 0.61–3.42). Local tobacco-related habits are not significant in this study. Alcohol consumption is also found to be a potential risk factor, but further study with larger population size is required to confirm it as statistically significant. Also some studies found that the risk of breast cancer was lower in women whose menarche occurred at age 15 or over, but there was no evidence for the risk to increase with decreasing age at menarche below age 15 years.

Sequence variation in ATPase and ND1 regions

Gene-specific PCR was performed in the 40 DNA (30 breast cancer and 10 healthy) samples isolated and specific products were observed. RFLP analysis was performed in all the samples and the digested products were documented. Thirteen (8 breast cancer and 5 healthy) samples were selected for sequencing based on the polymorphism observed in the gel picture. Sequence variations were found at 9 mtDNA ATPase (ATP6 and ATP8) at 9 different nucleotide positions in 5 samples (62.5 %) and a sequence variation of 2 mtDNA ND1 at two discrete nucleotide positions were found in 2 breast cancer samples (25 %). All the 11 variations from both ATPase and NDI were found to be base substitutions. Two novel tRNA-Leu mutation (8360A > T, 8362T > A) were observed for breast cancer in Mizo population. Most of the mutations were previously reported in open mutation mtDNA databases or in the literature (mtDBase:http://www.genpat.uu.se/mtDB/index.html; MITOMAP: http://mitomap.org/MITOMAP) and new mutations that are not reported in relationship with breast cancer were also found (Table 3). Sequence variations that are not recorded in the database were categorized as novel mutations, and those that appeared in the database were reported as polymorphisms. Among the 11 ATPase and ND1 mutations, 5 non-synonymous and 6 synonymous substitutions were found. Nine of these mutations were previously reported for other diseases and normal variations, but not for breast cancer in the literature or the unrestricted mtDNA mutation databases (mtDBase: http://www.genpat.uu.se/mtDB/index; MITOMAP:http://mitomap.org/ MITOMAP). In the ATPase 6 gene, 5 synonymous, and 3 non-synonymous mutations were observed and only 1 non-synonymous deleterious mutation was observed in the case of ATPase 8 gene. Furthermore, non-synonymous amino acid variants were more recurrent in the ATPase genes compared to the ND1 gene.

Evolutionary divergence of the cancer and healthy sequences

Genetic analyses were conducted on the three genes ATP6, ATP8, and ND1, separately. The relative values of nucleotide composition showed a moderate predominance of A and T over C and G in both cancer and healthy samples (Supplementary Fig. 1). Transition events were more frequent than transversions in all the three genes. The transitional status for all the three genes is higher than transversion status (Supplementary Fig. 2). The distance of divergence was found to be 0.003 for ATP6, 0.001 for ATP8, and 0.002 for ND1. Nucleotide substitutions per site are higher in ATP6 and ATP8 as compared to ND1. The log likelihood value for ATP6, ATP8, and ND1 genes were −962.808, −262.726, and −1284.541, respectively. The values of Tajima's D statistical test for ATP6, ATP8, and ND1 are −1.0433, −1.0062, and −1.2372, respectively. The distinctive features between the healthy and breast cancer samples in relationship with mitochondrial genes are shown in Table 4. The transition–transversion bias for the ATP6 is highest for cancer and healthy control sample. Also results from Tajima's Neutrality Test shows that the nucleotide diversity π, the haplotype frequency P, Watterson's Θ (Population mutation rate), which compares two estimators of the population parameter is higher in ATP6 compared to ATP8 and ND1. Thus, ATP6 may be a good marker for evolutionary variation based on the Tajima test statistic value (Table 4).

Discussion

Demographic studies earlier found that there are potential risk factors for breast cancer which includes older age, lack of breastfeeding, certain dietary patterns, obesity, and other reproductive factors such as age at menarche and menopause. The association between the risk of breast cancer and dietary fat intake has been controversial for decades. However, according to the findings of this study we found that high intake of animal fats, mainly from red meat during the premenopausal years are correlated with an

Table 3 Polymorphism and mutations in ATPase and ND1region of breast cancer samples

Gene name	Frequency of mutation (%)	Nomenclature of mutation	Codon number	Codon position	Syn/ nonsyn	Codon change	Amino acid change	Reported/ disease caused	Polyphen2 score
ATP6	12.5	8580 C > T	18	3	SYN	CCC > CCT	P > P	Reported but not for breast cancer	–
	12.5	8584G > A	20	1	NonSYN	GCA > ACA	A > T		0.011 (Benign)
	12.5	8598T > C	24	3	SYN	ATT > ATC	I > I	Reported for normal variation	–
	12.5	8602T > C	26	1	NonSYN	TTT > CTT	F > L	Reported but not for breast cancer.	0.001
	12.5	8701G > A	59	1	NonSYN	TTG > TTA	T > A	Reported for thyroid tumor. Not for breast cancer	(Benign)
	25	8790G > A	88	3	SYN	CTG > CTA	L > L	Reported but not for breast cancer	–
	12.5	8970C > T	148	3	SYN	AGC > AGT	S > S	Reported for normal variation	
	12.5	9090T > C	188	3	SYN	TCT > TCC	S > S		
ATP8	12.5	8414C > T	17	1	NonSYN	CTC > TTC	L > F	Reported for Alzheimer's disease and Parkinson's disease but not for breast cancer	0.994 Probably damaging (Novel report)
tRNA-Leu		8360A > T	–	–	–	–	–	Novel (Not reported in the database)	–
		8362T > A	–	–	–	–	–	Novel for breast cancer	
NDI	12.5	3505A > G	67	1	NonSYN	ACC > GCC	T > A	Reported for pancreatic and prostate cancer but not for breast cancer	0.002 (Benign)
	12.5	4065A > G	253	3	SYN	CAA > CAG	G > G	Reported but not for breast cancer	–

Table 4 Characteristic features of healthy and breast cancer samples

Gene Name	Transition/transversion bias (R)		Conserved region		Variable region		∏		Θ		S		Ps		D	
	H	B	H	B	H	B	H	B	H	B	H	B	H	B	H	B
ATP6	132.252	315.884	680	671	0	9	0.001468	0.004342	0.001602	0.005402	2	9	0.002937	0.013235	−0.709896	−1.043292
ATP8	0.391	293.289	207	206	0	1	0.000000	0.001380	0.00000E + 000	0.001972	0	1	0.000000	0.004831	n/c	−1.006231
ND1	140.134	139.259	954	954	2	2	0.001046	0.000598	0.001141	0.000854	2	2	0.002092	0.002092	−0.709896	−1.237160

S Number of segregating sites, *ps* S/n, *Θ* ps/a1, *π* nucleotide diversity, *D* Tajima test statistic, *H* Healthy, *B* Breast cancer patients

increased risk of breast cancer in Mizoram population. Most of the Mizo populations are non-vegetarians, and intake of animal fat is high in the form of pork meat and addition of fermented pork fat (saum) while preparing bai (vegetable stew) which is a prevalent dietary habit in this population. The mizo population is of mongoloid origin and are different from the rest of India in terms of their diet, lifestyle, and geographical distribution. Some studies have also established that countries with higher intake of animal fat such as dairy products and meat, are found to have higher breast cancer incidence [16–18]. Also in other mongoloids like Japanese, consumption of meat daily among women was found to increase the risk of breast cancer by 8.5 fold than women who never eat meat or seldom consume meat [19]. A collective investigation from several case–control studies concerning the correlations between total fat intake and the risk of breast cancer has suggested a positive association [20–22]. By contrast, prospective studies have not supported an association with total fat intake [23]. Population-based studies have established a fivefold difference in breast cancer occurrence between European countries and Asian countries, which suggests a role of diet and lifestyle in cancer prevalence.

Variations are also found in their incidence, mortality, and survival rates between different countries or population groups. Even various ethnic and racial groups are affected differently by overall cancer incidence. However, the reason for disparities is still not yet known, although recent research has suggested that genetic factors may be the reason for differences in breast cancer prognosis. Age at menopause is found to be a significant risk factor for this population, which may be correlated with longer exposure to female hormone estrogen. Also family history of other cancers are found to be associated with the increased risk of breast cancer in this population which may be because of their genetic make-up and inheritance of faulty genes, which makes them predispose to cancer. However, in the mizo population no significant association between breast feeding and risk of breast cancer was found, unlike other studies where positive correlations were found between breast feeding and risk for breast cancer. The epidemiological study contributes to the identification of several risk factors in relation with mitochondrial DNA for breast cancer given in the published international literature, with special reference to Mizo population [24, 25].

Also, results from our study revealed that ATP6 gene is more mutated as compared to ATP8 gene in breast cancer patients from Mizoram population as in other populations [26]. The number of occurrence of non-synonymous amino acid variants were also more often in the ATPase6 gene compared to the ATPase8 gene (Table 3). The G8584A, T8602C, and G8701A in ATPase6 gene lead to A20T, F26L, and T59A amino acid substitutions, respectively.

And the A3505G in ND1 gene leads to T67A amino acid substitution. Our results also indicate that the variants in ATPase6 gene are mostly benign which has no impact on protein structure as they were positioned in a poorly conserved region of protein based on polyphen-2 software. However, from polyphen-2 analysis, we found one novel mutation predicted as 'probably damaging variants' that leads to amino acid substitution of L (leucin) >F (phenylalanine) and the codon change from CTC > TTC at the 17th codon which may have an impact on the structure and function of ATPase8 gene which is located in highly conserved domains of the proteins. From the above result, we found 5 non-synonymous mutations which may have the potential to cause defects in the oxidative phosphorylation system by making the resulting protein to be non-functional or truncated. Also, out of the 8 breast cancer samples, two were most frequently found to be mutated in the ATP6 gene, which may be correlated with their epidemiological study as one of the patients was having a medical history of diabetes, hypertension and sciatica while the other had a nerve problem. In our study, we also found that the variations were more frequent in the ATPase region than in ND1 region of breast cancer patients, which is in contrast to previous studies which have suggested mtDNA ND1 as novel biomarker for the early detection of breast cancer. Also several authors have reported that variants of ATPase6 gene may increase the progression of cancer by inhibiting apoptosis pathways [27, 28]. Though the functional role of ATPase6 or ATPase8 variants in tumorigenesis is controversial; however, alteration in ATPase6 has been reported as a polymorphism in different studies. The rate of recurrence of polymorphism has been reported to be from 79 to 91.66 % in breast cancer patients [29, 30], and 75–100 % in other types of cancers [31–33]. The log likelihood is proportional to the probability of observing the data given the parameter estimates and our model. If the models are nested, then a larger likelihood vale means a larger probability of observing the data, which is goodness of fit. In our results, the log likelihood value for ATP6, ATP8, and ND1 are −962.808, −262.726, and −1284.541, respectively where ATP6 and ND1 are in Deviance and ATP8 is in goodness of fit. According to the Tajima statistics, $D < 0$ indicates that rare alleles are present at low frequencies; there was a recent selective sweep and an expansion of the genes after a recent bottleneck [34]. In our study, we found a strong selective sweep for ATP6 (D = −1.0433) and ND1 (D = −1.2372) genes for breast cancer samples based on the Tajima's D statistical test (Table 4). Hence, the high level of nucleotide diversity and Watterson's Θ (mutation rate) in ATPase genes belonging to breast cancer samples are essentially the large rate of total genetic variation in the mitochondrial genome.

Our study showed that ATPase genes are more predisposed to variation in breast cancer and may play an important role in tumorigenesis by changing the energy metabolism level in cancer cells. Also the diagnostic role of mtDNA for early detection of breast cancer may require further studies as to validate it as a potential biomarker. The present study is a novel finding in terms of the possible role of mtDNA ATPase and ND1 mutations in breast cancer in Mizoram population.

Acknowledgments This study was supported by by the State Biotech Hub (BT/04/NE/2009) sponsored by the Department of Biotechnology (DBT), New Delhi, Govt. of India.

Conflict of interest The authors report no conflict of interests. The authors alone are responsible for the content and writing of the paper.

References

1. Phukan RK, Zomawia E, Hazarika NC, Baruah D, Mahanta J. High prevalence of stomach cancer among the people of Mizoram, India. Curr Sci. 2004;87:3.
2. Perou CM, Sorlie T, Eisen MB, Van De Rijn M, Jeffrey SS. Molecular portraits of human breast tumours. Nature. 2000;406:747–52.
3. Man Yu, Yurong S, Fei Z, Yunli Z, Yi Y, Xiyin W, Lin Z, Ruifang N. Sequence variations of mitochondrial DNA D-loop region are highly frequent events in familial breast cancer. J Biomed Sci. 2008;15(4):535–43.
4. Yeole BB, Kurkure AP. An epidemiological assessment of increasing incidence and trends in breast cancer in Mumbai and other sites in India, during the last two decades. Asian Pac J Cancer Prev. 2003;4:51–6.
5. Yip CH, Taib NA, Mohamed I. Epidemiology of breast cancer in Malaysia. Asian Pac J Cancer Prev. 2006;7:369–74.
6. Saxena S, Kaushal M, Sharma J, Zomawia E, Kapur S. Genomic alterations in breast cancer patients from Northeast India using 10 K SNP arrays. Genome Biol. 2010;11:34.
7. Datta K, Biswas J. Influence of dietary habits, physical activity and affluence factors on breast cancer in East India—a case-control study. Asian Pac J Cancer Prev. 2009;10:219–22.
8. Cann RL, Stokening M, Wilson AC. Mitochondrial DNA and human evolution. Nature. 1987;325:32–6.
9. Anderson S, Bankier AT, Barrell BG, Bruijn MH, Coulson AR, Drouin J, et al. Sequence and organization of the human mitochondrial genome. Nature. 1981;290:457–65.
10. Czarnecka AM, Golik P, Bartnik E. Mitochondrial DNA mutations in human neoplasia. J Appl Genet. 2006;47:67–78.
11. Czarnecka AM, Marino Gammazza A, Di Felice V, Zummo G, Cappello F. Cancer as a 'Mitochondriopathy'. J Cancer Mol. 2007;3:71–9.
12. Czarnecka AM, Krawczyk T, Czarnecki JS, et al. Methodology for mitochondrial DNA research in oncology: goals and pitfalls. ARS Medica Tomitana. 2008;14:48–64.
13. Carew JS, Huang P. Mitochondrial defects in cancer. Molecular Cancer. 2002;1:1–12.
14. Jonckheere ISJAM, Rodenburg RJT. Mitochondrial ATP synthase: architecture, function and pathology. J Inherit Metab Dis. 2012;35:221–5.

15. Ghatak S, Muthukumaran RB, Nachimuthu SK. A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. J Biomol Tech. 2013;24(4):224–31.
16. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer. 1975;15:617–31.
17. Carroll KK, Braden LM. Dietary fat and mammary carcinogenesis. Nutr Cancer. 1985;6:254–9.
18. Rose DP, Boyar AP, Wynder EL. International comparisons of mortality rates for cancer of the breast, ovary, prostate, and colon, and per capita food consumption. Cancer. 1986;58:2363–71.
19. Hirayama T. Epidemiology of breast cancer with special reference to the role of diet. Prev Med. 1978;7:173–95.
20. Howe GR, Hirohata T, Hislop TG, Iscovich JM, Yuan JM, Katsouyanni K, et al. Dietary factors and risk of breast cancer: combined analysis of 12 case-control studies. J Natl Cancer Inst. 1990;82:561–9.
21. Armstrong B, Doll R. Environmental factors and cancer incidence and mortality in different countries, with special reference to dietary practices. Int J Cancer. 1975;15:617–31.
22. Boyd NF, Martin LJ, Noffel M, Lockwood GA, Tritchler DL. A meta analysis of studies of dietary fat and breast cancer risk. Br J Cancer. 1993;68:627–36.
23. Hunter DJ, Spiegelman D, Adami HO, Beeson L, van den Brandt PA, Folsom AR, et al. Cohort studies of fat intake and the risk of breast cancer—a pooled analysis. N Engl J Med. 1996;334:356–61.
24. Ghatak S, Doris L, Mawia L, Ricky S, Zothanpuia, Jeremy LP, Muthukumaran RB, Nachimuthu SK. Mitochondrial D-Loop and cytochrome oxidase subunit I polymorphisms among the breast cancer patients of Mizoram, Northeast India. Curr Genet. 2014;60:201–12.
25. Ghatak Lallawmzuali D, Mukherjee S, Mawia L, Pautu JL, Nachimuthu SK. Polymorphism in mtDNA control region of mizo-mongloid breast cancer samples as revealed by PCR-RFLP analysis. Mitochondrial DNA. 2014. doi:10.3109/19401736.2014.982627.
26. Massoud G, Reza M, Forouzandeh F, Behnam K, Nasrin M, Massoud H. The mitochondrial ATPase6 gene is more susceptible to mutation than the ATPase8 gene in breast cancer patients. Cancer cell international. 2014;14:21.
27. Shidara YYK, Kanamori T, Nakano K, Kwong JQ, Manfredi G, Oda H, Ohta S. Positive contribution of pathogenic mutations in the mitochondrial genome to the promotion of cancer by prevention from apoptosis. Cancer Res. 2005;65(5):1655–63.
28. Kirches E. Mitochondrial and nuclear genes of mitochondrial components in cancer. Curr Genomics. 2009;10(4):281–92.
29. Tan DJBR, Wong LJ. Comprehensive scanning of somatic mitochondrial DNA mutations in breast cancer. Cancer Res. 2002;62(4):972–6.
30. Czarnecka AMKA, Krawczyk T, Zdrozny M, Arnold RS, Bartnik E, Petros JA. Mitochondrial NADH-dehydrogenase polymorphisms as sporadic breast cancer risk factor. Oncol Rep. 2010;23(2):531–5.
31. Aikhionbare FOKM, Carey D, Okoli J, Go R. Is cumulative frequency of mitochondrial DNA variants a biomarker for colorectal tumor progression? Mol Cancer. 2004;3:30.
32. Aikhionbare FOMS, Kumaresan K, Zavareh M, Olatinwo M, Odunsi K, Partridge E. Mitochondrial DNA sequence variants in epithelial ovarian tumor subtypes and stages. J Carcinog. 2007;6:1.
33. Mehrabi SAJ, Adams G Jr, Grizzle W, Yao X, Aikhionbare FO. A sequence variant in colorectal adenomatous polyps. Diagn Pathol. 2010;5:66.
34. Tajima F. Statistical method for testing the neutral mutation hypothesis by DNA polymorphism. Genetics. 1989;123(3):585–95.

Short Communication

For reprint orders, please contact: reprints@futuremedicine.com

Personalized Medicine

Novel *AKT1* mutations associated with cell-cycle abnormalities in gastric carcinoma

Souvik Ghatak[‡,1], Sylvine Lalnunhlimi[‡,1], Freda Lalrohlui[1], Jeremy L Pautu[2], John Zohmingthanga[3], Ajaikumar B Kunnumakkara[4] & Nachimuthu Senthil Kumar*[,1]

[1]Department of Biotechnology, Mizoram University, Aizawl – 796004, Mizoram, India
[2]Mizoram State Cancer Institute, Zemabawk, Aizawl – 796017, Mizoram, India
[3]Department of Pathology, Civil Hospital Aizawl, Aizawl – 796001, Mizoram, India
[4]Department of Biotechnology, Indian Institute of Technology Guwahati, Guwahati – 781039, Assam, India
* Author for correspondence: Tel.: +91 943 635 2574; nskmzu@gmail.com
‡ Authors contributed equally and share the first authorship

Aim: The aim of this study is to identify the *AKT1* gene mutation driven pathogenicity in gastric cancer for Mizo population. **Methods:** 50 diffuse-type gastric tumors were analyzed for *AKT1* exon 2 and 14 mutations. Cell-cycle aberration was analyzed in the *AKT1*-mutated samples and the stability of the protein as well as exonic splicing enhancer motifs were examined. **Results:** The novel mutations, 15553T >A and 25376C >G might affect the exonic splicing enhancers and silencers. Significant decline was observed in the S-phase population in the tumor cells with 15553T >A and 15579G >C mutations suggesting the arrest of G1 phase. **Conclusion:** The present study is a novel finding of the possible role of *AKT1* mutations which might help to identify gastric cancer patients.

First draft submitted: 9 July 2017; Accepted for publication: 22 November 2017; Published online: 30 January 2018

Keywords: *AKT1* • cell cycle • exonic splicing enhancers • gastric cancer • pleckstrin homology domain

Several epidemiological risk factors that may be associated with gastric cancer have been identified [1]. AKT1 (also known as protein kinase B) plays a role in the regulation of cell survival, cell cycle, invasion and metabolism [2]. The homologous isoforms – AKT1, AKT2 and AKT*3* consist of three functional domains: an amino terminal pleckstrin homology (PH) domain, a central catalytic domain and carboxyl terminal regulatory domain with the hydrophobic motif. Numerous studies have shown aberrant activation of *AKT* through genetic events and its hyperactivation in human malignancies [3,4].

AKT is involved in important signaling pathways that comprises of *PI3K*, *PTEN* (phosphatase and tensin homolog) and mTOR maintaining a balance between cell survival and apoptosis [5]. The mTOR pathway frequently gets dysregulated in a variety of human cancers, including gastric cancer. Overexpression of the mTOR downstream effectors eIF4E and 4E-BP1 was found in gastric cancer cells [6]. The dysregulation of PI3K/PTEN/AKT/mTOR signaling pathway is a common event in several human cancers, including gastric cancer (GC) which could be triggered by loss of *PTEN* function, over expression or activating mutations of *AKT*, and over expression of upstream receptors such as *IGFR*, *EGFR* and *HER-2* [7].

So far, *AKT1* mutations have been reported in exons 2 and 14 of breast, colon, ovarian and lung cancers, and less information is available for gastric cancer [8]. Exon 2 codes for the PH domain and exon 14 for the regulatory hydrophobic motif. The PH domain recruits AKT to the plasma membrane by phosphoinositide binding and activates the AKT protein. The carboxyl-terminal hydrophobic regulatory domain contains several proline-rich regions that serves as protein–protein interaction sites and has significant role in regulation of AKT1 activity. The carboxyl-terminal hydrophobic regulatory domain contains F-X-X-F/Y-S/T-Y/F hydrophobic motif, where X can be any amino acid belonging to AGC kinase family. In mammals, this motif is identical (FPQFSY) in all the AKT isoforms and is very vital for the enzymatic activity [9]. Mutation in exon 2 alters the electrostatic interactions of the AKT protein pocket and forms new hydrogen bonds with a phosphoinositide ligand in PH domain and the

exon 14 is the mutational hot spot for the *AKT* gene and can alter the regulatory feature of AKT protein [8]. This study aims to assess the mutations in *AKT1* gene and cell-cycle alterations in gastric cancer patients.

Materials & methods

The mutations in exons 2 and 14 of the *AKT1* gene were analyzed from the formalin-fixed paraffin-embedded (FFPE) tumor and matched normal tissue of 50 diffuse-type gastric adenocarcinoma and compared with 50 healthy blood samples from Mizo, northeast Indian population. The samples consisted of 22 early and 38 advanced gastric carcinoma. Hematoxylin and eosin-stained slides were prepared after microdissection from tumor and matched normal tissue paraffin block, whereas the blood samples were directly used. Approval for this study was obtained from the Institutional Review Board of the Civil hospital, Mizoram as well as the Human Ethical Committee, Mizoram University, India.

Isolated genomic DNA from tumor tissue [10,11], matched normal tissue and blood were amplified with primer pairs covering the *AKT1* exon 2 (5′-CGCTGGCCCTAAGAAACA-3′ and 5′-CTCAGCTTTGGGACTCAGC-3′) and exon 14 (5′-TTAAGAGGTTGGCTTCCTACTG-3′ and 5′-CAGCGTGGCTTCTCTCAAAT-3′). 25 µl PCR was carried out, each containing 100 ng of template DNA, 0.2 pM of each primer, 2.5 µl of 10× PCR buffer, 1.5 mM $MgCl_2$, 200 mM dNTPs and 1 unit of Taq DNA polymerase (Fermentas, Sankt Leon-Rot, Germany). The reaction mixture was heated to 94°C for 5 min, followed by 30 cycles each consisting of 1-min denaturation at 94°C, 1-min annealing at 60°C (for both the exons) and 1 min of extension at 72°C and a final 5-min extension at 72°C. Single-strand confirmation polymorphism was performed for detection of mutation and the PCR products showing mobility shifts were sequenced.

14 PCR products were sequenced from the opposite direction to ensure reading accuracy. The chromatograms were examined using chromas software version 2.13 and aligned by Basic Local Alignment Search Tool (BLAST) [12]. To identify the exonic splicing enhancer motifs that were recognized by individual SR proteins, sequences were submitted to human splicing finder. Mutational identification and its effects on protein were found by Mutation Taster, PolyPhen2, Sorting Intolerant From Tolerant (SIFT) and Proven online server [13]. All annotated sequences were submitted to the European Bioinformatics Institute (EBI) repository server (accession numbers LN845722 – LN845747).

0.1 g of grossly gastric tumor and adjacent normal gastric mucosa tissues were used for cell-cycle analysis. Cells were harvested by mechanical dis-aggregation and fine-needle aspiration. Two separate aliquots of 6×10^6 tumor cells were prepared for each sample. Pellets were incubated with 250 ml of 0.1% RNAse (Sigma-Aldrich, MO, USA) and 50 mg/ml propidium iodide (presence of sodium citrate and TritonX-100) for 30 min at 37°C and flow cytometric analysis was performed. The proliferating cell population has four distinct phases: the G1-, S- (DNA synthesis phase), G2- and M-phase (mitosis). The G2- and M-phases will have identical DNA content and hence will not be discriminated [14].

To find the protein stability and the scores for free energy alterations for single-site mutation, the sequences were submitted to I-Mutant (version 2.0) [15] and FOLD-X energy-based web server. The FOLD-X tool gives the contrast between the wild- and mutant-type models in the form of van der Waals clashes, which deeply affect the energy breakdown [16,17]. NetSurfP server [18] was used to predict the surface accessibility and secondary structure of amino acids, based on Z-score [19]. Sequences were submitted to project Have yOur Protein Explained (HOPE [20]) for prediction of structural variation between wild- and mutant-type amino acids, which provides the 3D structural visualization of desire proteins and the results by using UniProt servers [19].

Results

We found mutations in four tumor samples and all were invasive diffuse-type carcinoma. Three sequence variations in exon 2 (in 8% samples) and two variations in exon 14 (in 12% samples) were observed (Figure 1). A 25376C >G homozygous mutation was very common in gastric cancer patients in exon 14 of the *AKT1* gene (Table 1). Due to this mutation, the splicing donor increased (wild-type: 0.39/mutant-type: 0.97) affecting the potential exonic splicing regulator and thereby leading to the loss of function of domain AGC-kinase C-terminal. In exon 2, splicing abnormalities were observed due to the mutation of 15553T >A. The PolyPhen2 and SIFT scores were pathological for these mutations. A 15553T >A mutation occurred in the late exonic position of the spliceosome recognition site (±40 nucleotide of splicing site). The splicing acceptor was increased (mutant-type: 80.58), hence cryptic splice site will get activated leading to new site generation at +56.07 site and decreasing the length of exon by 108 base pairs. Mutations of the 15553T >A and 25376C >G in two different exons are most likely to

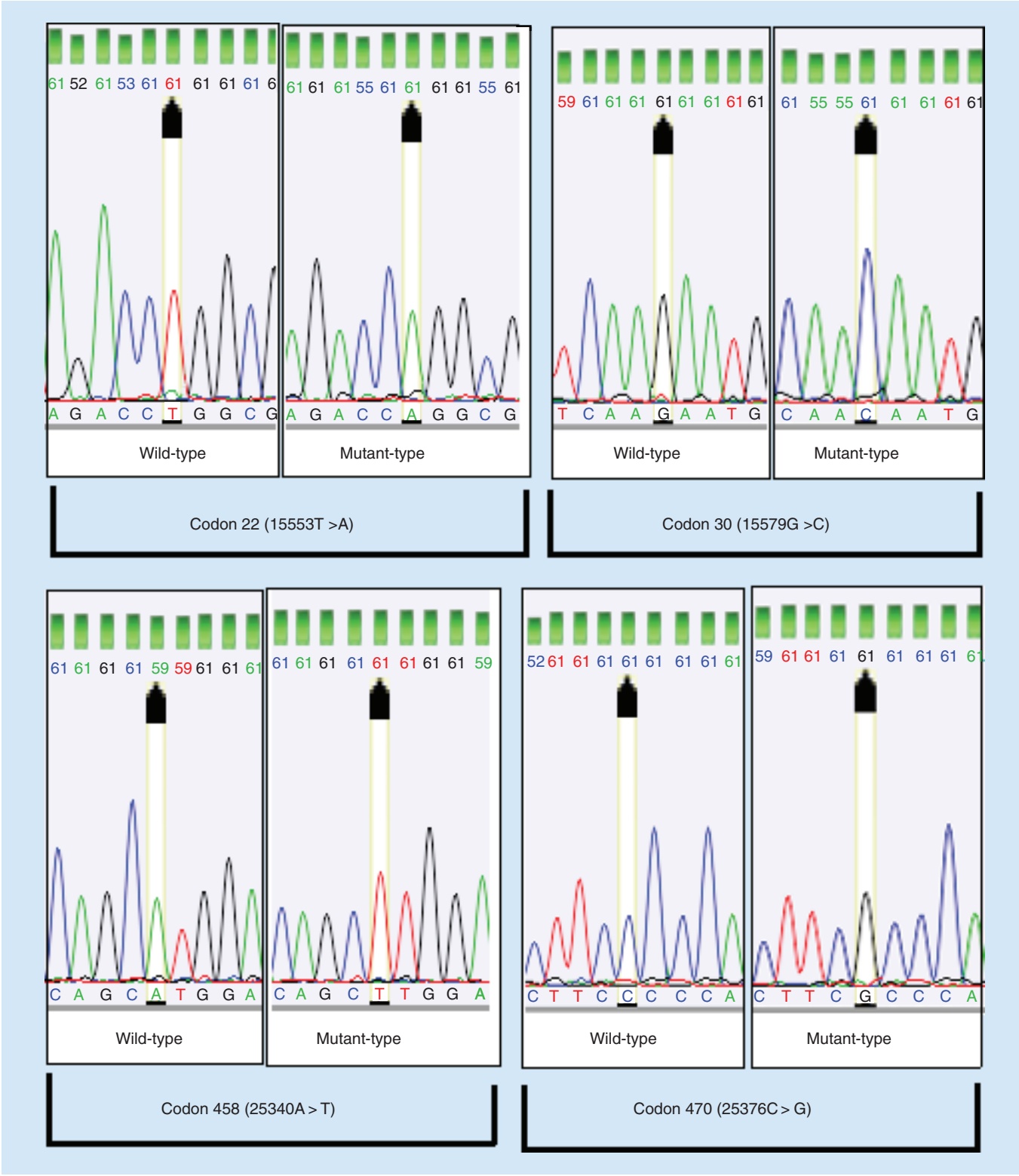


Figure 1. Chromatograms of the wild-type and mutant-type nucleotide sequences. The mutations are 15553T >A, 15579G >C, 25340A >T, 25376C >G.

Table 1. Mutations in *AKT1* gene and protein stability alteration in gastric cancer samples.

Position	Mutation name	Mutation frequency (%)	Codon changes	AA change	Polyphen2/SIFT/Proven	Stability	DDG (kcal/mol)	Class assignment basis of AA change		RSA/ASA	Effect of mutation by mutation taster
Exon 2	15553T > A[†]	6	TGG > AGG	W22R	(1.00) Damaging/damaging/deleterious	Decrease	-0.84	W	Exposed	0.0354/85.233	Effect on PHD function. Potential splice site changes found. Acceptor increased (wt: 51.63/mu: 80.58)
								R	Buried	0.252/57.800	
	15579G > C[†]	8	AAG > AAC	K30N	(0.395) Benign/tolerated/deleterious	Decrease	-1.85	K	Exposed	0.296/60.846	Loss of PHD function
								N	Exposed	0.257/37.698	
	15645C > T	8	CTC > CTT	L42L	Natural	–	–	–	–	–	Not effective
Exon 14	25340A > T[†]	10	ATG > TTG	M458L	(0.00) Benign/tolerated/natural	Decrease	-0.50	M	Exposed	0.382/76.478	Loss of domain AGC-kinase C-terminal function
								L	Exposed	0.325/59.526	
	25376C > G[†]	12	CCC > GCC	P470A	(0.430) Probably damaging/tolerated/deleterious	Decrease	-0.47	P	Exposed	0.560/79.492	Loss of domain AGC-kinase C-terminal function and donor increased (wt: 0.39/mu: 0.97)
								A	Exposed	0.559/61.613	

[†] Represents novel mutation.
ASA: Absolute surface accessibility; DDG: DDG is the stability (DDG < 0: Decreased stability, DDG > 0: Increased stability); mu: Mutant; PHD: Pleckstrin homology domain; RSA: Relative surface accessibility, wt: Wild-type.

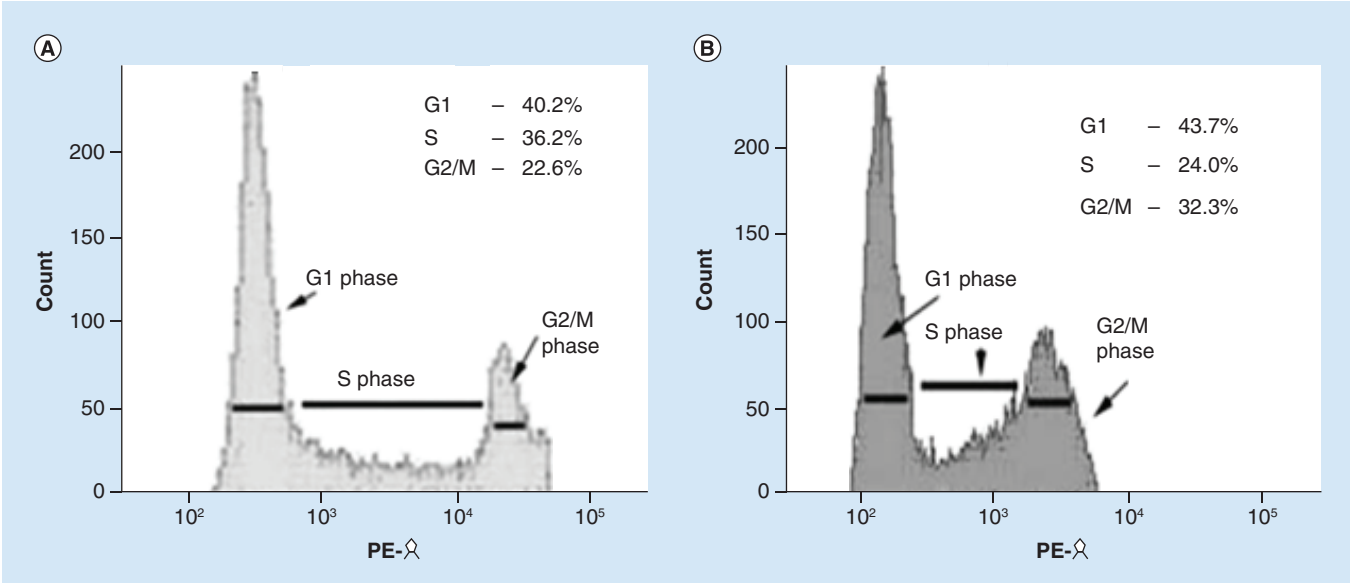


Figure 2. Cell-cycle histogram for (A) matched normal and (B) gastric cancer tumor cell.

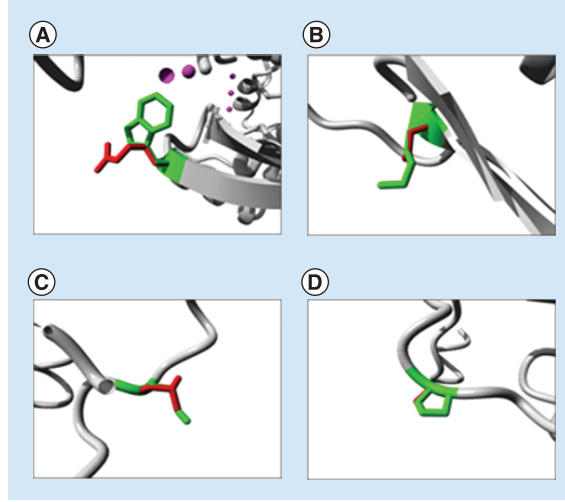


Figure 3. 3D structure of the mutation in gastric cancer samples. Both the wild-type and mutant side chain are shown in green and red, respectively. The rest of the protein is shown in grey. **(A)** Mutation of tryptophan to arginine at position 22 due to 15553T >A; **(B)** mutation of lysine to asparagine at position 30 due to 15579G >C; **(C)** mutation of methionine to leucine at position 458 due to 25340A >T; **(D)** mutation of proline to alanine at position 470 due to 25376C >G.

affect exon splicing, due to alteration of exonic splicing enhancers and silencers and denoted as exonic splicing regulator-relevant mutations (Table 1).

Significant elevation in DNA quantity in G1 phase and decrease in the S-phase population in tumor cells were observed (Figure 2A & B). Remarkably, the results were found in two samples with 15553T >A and 15579G >C mutations suggesting the arrest of G1 phase and entering into S phase due to the alteration of PH domain. Hence, loss of the protein PH domain can also alter the cell cycle and increase the tumor growth due to the mutations.

In this study, P470A mutation occurred frequently (in 12% samples) than any other mutation which was giving a probable damage score in SIFT and polyphen2. This mutation also might increase Akt phosphorylation and the AKT active conformation might also change. The imino acid proline is converted to nonpolar alanine resulting in loss of domain AGC-kinase C-terminal function and increase in splicing donor region (wild-type: 0.39/mutant: 0.97; Figure 3D).

There is a significant alteration in charge of amino acid (22W >R) between the mutant and wild-type due to 15553T >A substitution (Table 1). The mutation brings in positive charge at this position; causing repulsion between the mutants and neighboring residues. The mutant has a smaller residue than the wild-type (Figure 3A), which will cause a failure of external interactions. Arg22 mutation induces changes in the conformation and protein

stability which may be due to clashes with electronic density of vicinity residues. The mutation might cause loss of hydrophobic interactions with other molecules on the surface of the protein. The mutation 458M >L might cause an empty space in the core of the protein due to the mutant residue being smaller than the wild-type (Figure 3C). Most of the mutant residues showed low relative surface accessibility and absolute surface accessibility (Table 1), but the mutation of tryptophan to arginine showed increase in relative surface accessibility. The exposed tryptophan large residue is converted into arginine in the buried small moiety form (Figure 3A).

Discussion

In the present study, novel mutations in gastric cancer have been identified in the *AKT1* gene. In Mizo population, most of the reported gastric cancers are diffuse-type and the tumors are present in the upper third of the stomach. Regarding diffuse-type gastric cancer, patients with mutations in the PI3K/AKT pathway are more likely to have tumors in the upper third of the stomach [21]. In exon 2, splicing abnormalities were observed due to the mutation (15553T >A). The PH domain function might be lost which holds AKT in a closed arrangement under basal conformation [22], due to the 15553T >A and 15579G >C mutations. Further, due to this mutation, exonic cryptic acceptor site in the presence of one or more cryptic branch points may be activated, hence potential alteration of splicing can occur [23]. PH domain contains leucine-rich repeat protein phosphatase which can induce apoptosis and inhibit tumor growth by dephosphorylation of S473 on AKT [24].

It is also possible that as in prostrate and colorectal cancers, mutated *AKT* may deregulate cell-cycle control and enhance survival after cytotoxic exposure [25,26]. AKT is involved in the control of G2/M cell-cycle progression and activation of AKT can overcome both the p53-independent G2/M cell-cycle checkpoint and apoptosis induced by DNA damage [26]. In addition, activation of AKT also has the ability to alleviate the p53-mediated cell-cycle checkpoints through phosphorylation and better degradation of p53 [27]. In our study, the samples with 15553T >A and 15579G >C mutations suggest the arrest of G1 phase entering into S phase due to the alteration of PH domain and degradation of p53. Aromatic tryptophan to polar arginine substitution at 22 amino acid position might affect the function of intracellular signaling due to PH domain alteration and resulting in frequent change in protein active or binding site alteration.

In this study, P470A mutation caused AKT1 amino acid tail deletion leading to higher AKT phosphorylation signifying that either phosphorylation or deletion of the tail region could lock AKT in its active conformation [28]. Phosphorylation of the AKT1 tail activated AKT1 leads to elevated Skp2 or FOXO phosphorylation in turn leading to elevated cell-cycle progression [27].

By studying the positional changes and interactions of amino acid residues within AKT1 protein, a significant number of structural changes have been identified that can explain the activity deviations caused by several mutations which were described previously by Almhanna *et al.* [4]. The PI3K/AKT pathway is a prototypic survival signal that is constitutively activated due to *AKT* gene amplification or mutations in *AKT* gene signaling components [29]. Once activated, AKT signals are propagated to diverse pathways which are attractive drug targets in cancer and have also been reported to confer resistance to cancer therapies leading to poor prognosis for several tumors [29]. In exon 2, the PH domain function might be lost due to splicing abnormalities. PH domain holds AKT in a closed arrangement under basal conformation, which recruits AKT to the plasma membrane by phosphoinositides binding and this binding is required for activation of the AKT protein for normal regulation of P13/AKT pathway.

In this study, the limitations concerning sample size may be overlooked because of the fact that the Mizo ethnic group has a very small population size of about 1.1 million people [30]. The fact that our study is one of the first performed in the unique and high gastric cancer risk Mizo-Mongoloid population from northeast India will provide insights for early diagnosis of this critical disease.

Conclusion

In conclusion, we have identified four nonsynonymous and one synonymous mutations in exons 2 and 14 of the *AKT1* gene in gastric cancer samples. The mutations were somatic as the *AKT1* exon variations were derived from tumor samples and no variation was observed in matched normal tissues. The present study is a novel finding in terms of AKT1 mutations which might help to identify patients who are at high risk for gastric cancer.

Acknowledgements

The authors thank A Bhaumik, Pathologist, Agartala Govt Medical College, Tripura for his permission and help in using the flow cytometer and to DK Zorinsanga, Department of Biotechnology, Mizoram University for his help during the work.

Executive summary

- This study has identified four nonsynonymous and one synonymous mutations in *AKT1* exons 2 and 14 in gastric cancer Mizo population.
- The mutations are somatic as the *AKT1* exon variations were derived from tumor samples and no variation was observed in matched normal tissues.
- Significant decline was observed in the S-phase population in the tumor cells with 15553T >A and 15579G >C mutations suggesting the arrest of G1 phase.
- The pleckstrin homology domain function might be lost which holds AKT in a closed arrangement under basal conformation, due to the 15553T >A and 15579G >C mutations.
- In exon 2, the pleckstrin homology domain function might be lost due to splicing abnormalities.
- P470A mutation caused AKT1 amino acid tail deletion leading to higher AKT phosphorylation signifying that either phosphorylation or deletion of the tail region could lock AKT in its active conformation.
- The present study is a novel finding in terms of the possible role of *AKT1* mutations which might help to identify patients from other populations as well who are at high risk for gastric cancer.
- The mentioned *AKT*1 mutation and cell-cycle abnormality can be used for early diagnosis marker for high risk of gastric cancer.

Financial & competing interests disclosure

This work was supported by Department of Biotechnology, Govt of India, New Delhi through the Advanced State Biotech Hub (BT/04/NE/2009 dt. 29.08.2014). The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

Approval for this study was obtained from the Institutional Review Board of the Civil Hospital, Aizawl and Mizoram University, Aizawl, Mizoram, India.

Informed consent disclosure

The authors state that they have obtained verbal and written informed consent from the patient/patients for the inclusion of their medical and treatment history within this case report.

References

Papers of special note have been highlighted as: • of interest; •• of considerable interest

1. Malakar M, Devi KR, Phukan RK *et al.* p53 codon 72 polymorphism interactions with dietary and tobacco related habits and risk of stomach cancer in Mizoram, India. *Asian Pac. J. Cancer Prev.* 15, 717–723 (2014).
2. Gschwind A, Fischer OM, Ullrich A. The discovery of receptor tyrosine kinases: targets for cancer therapy. *Nat. Rev. Cancer* 4, 361–370 (2004).

•• **A comprehensive review of the literature demonstrating about the receptor tyrosine kinase PI3K and the survival mediator AKT and their effect of cancer.**

3. Davies MA. Regulation, role and targeting of AKT in cancer. *J. Clin. Oncol.* 29, 4715–4717 (2011).
4. Almhanna K, Strosberg J, Malafa M. Targeting AKT protein kinase in gastric cancer. *Anticancer Res.* 31, 4387–4392 (1994).

• **Demonstrates that AKT plays an important part in cell-cycle regulation, is critical to gastric cancer cell survival and proliferation, and plays a significant role in chemotherapy resistance which makes it an attractive target for anticancer therapy.**

5. Boormans JL, Korsten H, Ziel-vander ACJ, van-Leenders GJLH, Verhagen PCMS, Trapman J. E17K substitution in AKT1 in prostate cancer. *Br. J. Cancer* 102, 1491–1494 (2010).
6. Qiu MZ, Xu RH. The progress of targeted therapy in advanced gastric cancer. *Biomark. Res.* 32(1), 1–8 (2013).
7. Wong H, Tau T. Targeted therapy in the management of advanced gastric cancer: are we making progress in the era of personalized medicine? *Oncologist* 17, 346–358 (2012).

•• **The conclusion of the study is that the targeted therapy for advanced gastric cancer patients can beimproved by identifying biomarkers involved in gastric carcinogenesis indifferent geographical populations.**

8. Malanga D, Scrima M, De-Marco C *et al.* Activating E17K mutation in the gene encoding the protein kinase AKT in a subset of squamous cell carcinoma of the lung. *Cell Cycle* 5, 665–669 (2008).

- **Demonstrated that E17K mutation increases AKT1 activity and localization in tumor cells, which will affect cell-cycle regulation.**

9. Jiang T, Qiu Y. Interaction between Src and a C-terminal proline-rich motif of Akt is required for Akt activation. *J. Biol. Chem.* 278(18), 15789–15793 (2003).

- •• **Suggests that interaction between the SH3 domain of Src family kinases and the proline-rich motif in the C-terminal regulatory region of Akt is required for tyrosine phosphorylation of Akt and its subsequent activation.**

10. Ghatak S, Muthukumaran RB, SenthilKumar N. A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. *J. Biomol. Tech.* 24, 224–231 (2013).
11. Ghatak S, Sanga Z, Pautu JL, Kumar NS. Co-extraction and PCR based analysis of nucleic acids from formalin-fixed paraffin-embedded specimens. *J. Clin. Lab. Anal.* 29(6), 485–492 (2015).
12. NIH. BLAST®. www.ncbi.nlm.nih.gov/blast
13. Ghatak S, Doris L, Mawia L *et al.* Mitochondrial D-Loop and cytochrome oxidase subunit I polymorphisms among the breast cancer patients of Mizoram, northeast India. *Curr. Genet.* 60, 201–212 (2014).
14. Blanco R, Rengifo CE, Cedeño M, Frómeta M, Rengifo E. Flow cytometric measurement of aneuploid DNA content correlates with high S-phase fraction and poor prognosis in patients with non-small-cell lung cancer. *ISRN. Biomarkers* 2013, 1–8 (2013).
15. I-Mutant (version 2.0). http://folding.uib.es/i-mutant/i-mutant2.0.html
16. Abagyan R, Totrov M. Biased probability Monte Carlo conformational searches and electrostatic calculations for peptides and proteins. *J. Mol. Biol.* 235, 983–1002 (1994).
17. Schymkowitz J, Borg J, Stricher F, Nys R, Rousseau F, Serrano L. The FoldX web server: an online force field. *Nucleic Acids Res.* 33, W382–W388 (2005).
18. NetSurfP. www.cbs.dtu.dk/services/NetSurfP/
19. Petersen B, Petersen TN, Andersen P, Nielsen M, Lundegaard C. A generic method for assignment of reliability scores applied to solvent accessibility predictions. *BMC Struct. Biol.* 9, 51–61 (2009).
20. HOPE. www.cmbi.ru.nl/hope/
21. Fang WL, Huang KH, Lan YT *et al.* Mutations in PI3K/AKT pathway genes and amplifications of *PIK3CA* are associated with patterns of recurrence in gastric cancers. *Oncotarget* 7(5), 6201–6220 (2016).

- •• **Patients with P13K/AKT pathway alteration were more likely to have diffuse-type and poorly differentiated gastric cancers and were more likely to experience peritoneal recurrence.**

22. Venselaar H, Te-Beek TA, Kuipers RK, Hekkelman ML, Vriend G. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. *BMC Bioinformatics* 11, 548 (2010).
23. Wu WI, Voegtli WC, Sturgis HL, Dizon FP, Vigers GPA, Brandhuber BJ. Crystal structure of human AKT1 with an allosteric inhibitor reveals a new mode of kinase inhibition. *PLoS ONE* 5(9), e12913 (2010).
24. Zhu M, Chen HM, Wang YP. Missense mutations of MLH1 and MSH2 genes detected in patients with gastrointestinal cancer are associated with exonic splicing enhancers and silencers. *Oncol. Lett.* 5, 1710–1718 (2013).
25. Gao T, Furnari F, Newton AC. PHLPP: a phosphatase that directly dephosphorylates Akt, promotes apoptosis, and suppresses tumor growth. *Mol. Cell* 18, 13–24 (2005).
26. El-Deiry WS. Akt takes centre stage in cell-cycle deregulation. *Nat. Cell Biol.* 3, E71–E73 (2001).

- **Described how knowledge of the Her2/neu-AKT-p21-Cip1/WAF1 pathway may be of use in designing more effective cytotoxic therapeutic strategies for Her2/neu-overexpressing, AKT-mutated or PTEN-mutated cancers.**

27. Kandel ES, Skeen J, Majewski N *et al.* Activation of Akt/protein kinase B overcomes a G2/M cell cycle checkpoint induced by DNA damage. *Mol. Cell Biol.* 22(22), 7831–7841 (2002).
28. Li J, Dang Y, Gao J *et al.* PI3K/AKT/mTOR pathway is activated after imatinib secondary resistance in gastrointestinal stromal tumours (GISTs). *Med. Oncol.* 32, 111 (2015).

- **Demonstrated that the PI3K/AKT/mTOR pathway can be partly activated after imatinib secondary resistance in gastrointestinal stromal tumors.**

29. Lee BL, Lee HS, Jung J *et al.* Nuclear factor-κB activation correlates with better prognosis and Akt activation in human gastric cancer. *Clin. Cancer Res.* 11(7), 2518–2525 (2005).
30. Anon. Chapter 2: size, growth rate and rural-urban distribution of populations in Mizoram, provisional populations totals – Mizoram. India: Census of India, 7–13 (2011).

Helicobacter

Helicobacter ISSN 1523-5378
doi: 10.1111/hel.12308

Xenobiotic Pathway Gene Polymorphisms Associated with Gastric Cancer in High Risk Mizo-Mongoloid Population, Northeast India

Souvik Ghatak,* Ravi Prakash Yadav,* Freda Lalrohlui,* Payel Chakraborty,* Soumee Ghosh,† Sudakshina Ghosh,† Madhusudan Das,† Jeremy L. Pautu,‡ John Zohmingthanga§ and Nachimuthu Senthil Kumar*

*Department of Biotechnology, Mizoram University, Aizawl, Mizoram, India, †Department of Zoology, University of Calcutta, Kolkata, West Bengal, India, ‡Mizoram State Cancer Institute, Aizawl, Mizoram, India, §Department of Pathology, Civil Hospital, Aizawl, Mizoram, India

Keywords
Gastric cancer, *Helicobacter pylori*, *CagA*, *glutathione S-transferases*, polymorphisms, Mizo-Mongoloid.

Reprint requests to: Nachimuthu Senthil Kumar, Department of Biotechnology, Mizoram University, Aizawl - 796 004, Mizoram, India.
E-mail: nskmzu@gmail.com.

Abstract

Background: The aim of this study was to evaluate the risk of gastric cancer associated with individual or combined glutathione S-transferases (GSTs) polymorphism and their interaction with environmental factors.

Materials and Methods: Genotyping by PCR was carried out for 80 cases and controls each for *GSTM1*, *GSTT1,* and *GSTP1* polymorphism and mapped for gene—environment association studies. The samples were subjected to pathogen detection and *GSTP1* expression for analyzing their association with different genotypes. Logistic regression analyses were conducted to compute the influence of both genetic and environmental factors for gastric cancer. MDR analysis was performed to assess the risk of gastric cancer by studying the gene–gene and gene–environment effect on the basis of GST genotyping and *GSTP1* gene expression.

Results: Infection with *Helicobacter pylori* and *CagA*+ strains was more frequent in patients with GSTM1/T1 null genotype. Intake of high fermented fat and smoked meat was found to be significantly associated with gastric cancer. The G/G, A/G (rs1695), and T/T (rs1138272) were found to be significantly associated with low expression of GSTP1 gene in cancer tissue.

Conclusion: Presence of *H. pylori* with *CagA* genotype showed significant individual effect with *GSTT1* polymorphism as well as strong synergistic effect in gastric cancer risk. Majority of the gastric cancer samples showed significant negative expression in G/G, A/G (rs1695), and T/T (rs1138272) genotypes. This study shows that *GST* gene polymorphism was significantly relevant for determining the individual susceptibility to gastric cancer.

Stomach cancer is the second leading cause of death and the fourth most common malignancy in the modern world and highly associated with dietary and lifestyle habits [1,2]. Considerable differences in the incidence of stomach cancer were found worldwide, and the highest incidence among men has been reported from China (AAR = 145.0) and among women from Japan (AAR = 38.9) per 10^5 population [2]. It is generally accepted that cancer risk is influenced by the interaction between genetic components and environmental factors, such as diet and lifestyle. Previous studies on risk factors for stomach cancer have shown a positive correlation between *GST* polymorphism and consumption of alcohol, tobacco smoking, unique food habits, and pathogen infection [3,4]. Thus, the studies of genes involved in the detoxification of environmental carcinogens are important for the determination of interindividual susceptibility to cancer [5].

Stomach cancer is a multifactorial disease and arises from the interactive combinations of allelic variants (in low-penetrance genes) and relevant environmental risk factors. Genetic polymorphism in xenobiotic metabolizing genes may act as susceptibility factors for stomach cancer risk. Glutathione S-transferases (*GST*s) constitute a superfamily of phase-II metabolizing enzymes that play a key function in detoxification of a wide variety of potentially cytotoxic/genotoxic compounds [6]. The *GST* isoenzymes (*GSTM-1*, *GSTT-1*, and *GSTP-1*) are normally expressed along the human gastrointestinal tract [7]. The phenotypic absence of *GSTM1* and *GSTT1* activity is due

to homozygosity for an inherited deletion of these genes, termed as null genotype [8]. Individuals with the homozygous *GSTM1* null genotype show no protein expression and are expected to have reduced detoxification ability for hazardous compounds [9]. Furthermore, single-nucleotide polymorphisms (SNPs) in the *GSTP1* gene at codons 105 (Ile→Val) and 114 (Val→Ala) were found to be associated with reduced GST enzyme activity for several classes of substrates [10–13].

However, the expression of *GSTP1* influenced by polymorphism and gene–gene and gene–environment interactions among tobacco-associated patients with gastric cancer is not well understood. An individual's difference in susceptibility to chemically induced carcinomas may possibly be attributed to the variations in the detoxification pathways due to polymorphic variants of GST genes [13]. In European population, the *GSTM1* deletion frequency was found to be higher than *GSTT1* [13], and a significant association was found between the *GSTT1* null genotype and GC risk with all other *GST* isoenzymes polymorphism [14]. According to previous studies, when stratified by ethnicity, a significant association between *GSTT1* null genotype and GC risk was also found in Chinese and other Asian populations, but not in Caucasians [14].

Mizoram, northeast India, records the highest incidence of gastric cancer in India [15]. Consumption of tobacco in various traditional forms such as "tuibur" (tobacco smoke–infused aqueous solution) and Chaini (traditional name "Sada") is common habit that is important risk factor for stomach cancer [15,16]. Therefore, correlation of consumption of tobacco with the incidence of stomach cancer in Mizoram cannot be ruled out. We performed a case–control study to evaluate the relevance of *GSTM1*, *GSTT1,* and *GSTP1* gene polymorphisms to gastric cancer (GC) susceptibility and to assess their interaction with other environmental and lifestyle factors, namely *H. pylori*, *EBV* infection, and smoking habits for this high-GC risk state of Mizoram, northeast India.

Materials and Methods

Subjects

High incidence of stomach cancer is found in the Mizo-Mongoloid population, northeast India, with age-adjusted rates (AAR) of 42.9 and 20.5 per 10^5 population in male and female, respectively [17]. This study consisted of 80 random patients with gastric cancer (aged from 37 to 79 years) with or without a family history of any cancer and 80 unrelated healthy controls (aged from 31 to 73 years) belonging to the same ethnic group. The samples were collected from patients who received treatment from Mizoram State Cancer Institute, Zemabawk, Mizoram, India, between September 2013 and February 2015. The study design and data collection methods have been described previously. Patients with gastric neoplasms including adenocarcinoma (MALT lymphoma, stromal, or carcinoid tumors) were included. Patients with secondary or recurrent GC, previous history of other malignancies, or refusal to participate were excluded from the study [18]. Medical charts were reviewed using a standard protocol to obtain information on cancer treatment, clinical stage, dietary habits, previous disease history, physical activity, tobacco, and alcohol use by an in-person interview using a structured questionnaire [18]. The study protocol was approved by the Institutional Review Board of all institutes involved in the study.

Nucleic Acid Extraction

Blood samples (50 μL) were processed freshly and served as subjects for DNA isolation [19]. Hematoxylin and eosin-stained slides were prepared after microdissection from paraffin-embedded tumor tissue for conforming the adenocarcinoma, whereas the blood samples were used directly. Coextraction of RNA and DNA was performed for the paraffin-embedded tumor tissue [20].

Genotyping of GST M1/T1/P1 Gene Polymorphisms

GSTM1 and *GSTT1* null polymorphisms were detected by multiplex polymerase chain reaction (PCR) (Table 1) [21]. The reaction mixture (25 μL) contained 50–100 ng of genomic DNA in 1X Taq buffer, 200 μmol/L of each dNTP, 0.15 μmol/L of each primer, and 1 U of Taq DNA polymerase. Amplified products were analyzed by electrophoresis on 8% polyacrylamide gels resulting in a 219-bp fragment for *GSTM1*, 459-bp fragment for *GSTT1,* and a 349-bp fragment of the albumin gene (as an internal control). The absence of the specific *GSTM1* and/or *GSTT1* fragments indicated the corresponding null genotype, whereas the presence of the 349 bp albumin band ensured that the null genotype was not documented due to failure of PCR. In addition, two SNPs in the *GSTP1* gene for amino acid substitution at codons 105 (Ile→Val) and 114 (Val→Ala) were genotyped by PCR-RFLP method (Table 1) [11,22,23]. The fragment containing the *GSTP1* Ile105-Val (rs1695) polymorphic site was amplified according to the following parameters: 94 °C for 5 minutes, followed by 35 cycles at 94 °C for 30 seconds, 58 °C for 30 seconds, 72 °C for 1 minutes, and a final elongation at 72 °C for 7 minutes. Ten μL of PCR product was

Table 1 Primers tested for polymorphism pattern in the present study

Gene	Primer (5′ to 3′)	Annealing temperature	Type of polymorphism	Amplified products/Alleles
GSTM1	F: GAACTCCCTGAAAAGCTAAAGC R: GTTGGGCTCAAATATACGGTGG	60 °C–1 minutes	Gene deletion	– Null genotype+ 219 bp
GSTT1	F: TTCCTTACTGGTCCTCACATCTC R: TCACCGGATCATGGCCAGCA		Gene deletion	– Null genotype+ 459 bp
Albumin	F: GCCCTCTGCTAACAAGTCCTAC R: CCCTAAAAAGAAAATCGCCAATC		–	Internal control 349 bp
GSTP1 I105V (rs1695)	F: AATACCATCCTGCGTCACCT R: TGAGGGCACAAGAAGCCCCTT	59 °C–40 seconds	BsmAI RFLP	A - Digested fragment (308, 258 bp) G - Digested fragment (258, 219 and 89 bp)
GSTP1V114A (rs1138272)	F: ACAGGATTTGGTACTAGCCT R: AGTGCCTTCACATAGTCATCCTTG	50 °C–40 seconds	AciI RFLP	C - Digested fragment (143, 27 bp) T - Intact fragment (170 bp)
H. pylori (16srRNA)	F: CTGGAGAGACTAAGCCCTCC R: ATTACTGACGCTGATTGTGC	60 °C–45 seconds	Multiplex PCR	109 bp
H. pylori (CagA)	F: AATACACCAACGCCTCCAAG R: TTGTTGCCGCTTTTGCTCTC			400 bp
EBNA3C	F: AGAAGGGGAGCGTGTGTTGT R: GGCTCGTTTTTGACGTCGGC	56 °C–45 seconds		153 bp for type 1 EBV 246 bp for type 2 EBV
GSTP1 For expression	F: ATGACTATGTGAAGGCACTG R: AGGTTCACGTACTCAGGGGA	55 °C–45 seconds	Multiplex PCR	433 bp band (Intensity less—less expiration; Intensity more—optimum expiration)
β-Actin Control	F: ACCATGGATGATGATATCGC R: ACATGGCTGGGGTGTTGAAG			392 bp band Band Intensity should be equal

digested with 1 unit of *BsmAI* (New England Biolabs, Barcelona, Spain) for 6 hour at 55 °C. Digests were electrophoresed on 8% polyacrylamide gels resulting in three fragments of 305, 135, and 128 bp (allele A) or in four fragments of 222, 135, 128, and 83 bp (allele G). PCR cycling conditions were 5 minutes at 94 °C followed by 35 cycles at 94 °C for 30 seconds, 52 °C for 30 seconds, and 72 °C for 30 seconds with a final elongation step at 72 °C for 7 minutes for *GSTP1* Val114Ala (rs1138272) polymorphism. Ten μL of PCR products was digested with 1 unit of *AciI* (New England Biolabs, Barcelona, Spain) for 6 hour at 37 °C and electrophoresed on 8% polyacrylamide gel. The T allele was defined by the presence of an intact fragment of 170 bp and the C allele by the presence of two fragments of 143 and 27 bp.

PCR-Based Pathogen Detection

The presence of *H. pylori* infection was determined in patients with GC by multiplex PCR amplification of *16S rRNA* and *CagA* genes using Hp1-Hp2 and CagAF-CagAR primers, respectively (Table 1) [20,24]. The PCR products were electrophoresed with 1.5% agarose gel. *H. pylori* infection was defined by the presence of an intact band of 109 bp (16S rRNA) and 400 bp (*CagA* gene). The presence of *EBV* type1/type 2 infections was carried out by a standard PCR assay crossways type-specific regions of EBNA3C gene using previously described primer sets (Table 1) [25]. The PCR amplification was carried out for 35 cycles at 95 °C for 45 seconds, 56 °C for 45 seconds, 72 °C for 1 minutes followed by a final extension at 72 °C for 10 minutes. The PCR products were analyzed by 12% polyacrylamide gel, and the presence of 153 and 246 bp was considered as type 1 and type 2 EBV, respectively.

Reverse Transcriptase PCR (RT-PCR) for *GSTP1* Expression

First-strand synthesis of cDNA was performed using Superscript II reverse-transcriptase (Life Technologies Inc., USA) (Table 1). Five μg aliquot of total cellular RNA was used for each reverse transcription reaction, and one-tenth of this reaction was used for PCR. Primers targeting exons 6 and 7 of *GSTP1* and exons 1 and 3 of β-actin were used [13]. The PCR for *GSTP1* was run for 38 cycles consisting of 94 °C for 30 seconds, 1 minutes at gradually decreasing temperatures using a touchdown protocol (4 cycles each at 62 °C and 60 °C; and 30 cycles at 58 °C), and 74 °C for 5 minutes. RT-PCR products

were electrophoresed in 10% polyacrylamide gel and visualized by ethidium bromide staining.

Densitometric Analysis

Semiquantitative RT-PCR gels were analyzed to compare the banding pattern and their molecular mass using gel image analysis software Syngen G-Box (Sacramento, CA, USA) and ImageJ. The *GSTP1* expression was compared with *GSTP1* genotyping by HEMI 1.0.1-Heatmap Illustrator software.

Statistical Analysis

GST gene polymorphism among case–control subjects was tested for Hardy–Weinberg equilibrium by a chi-square ($\chi 2$) test with one degree of freedom (df). The polymorphisms in each group were estimated using odds ratios (ORs), and 95% confidence intervals (CIs) for each genotyping. *GSTT1* and *GSTM1* null genotyping were checked by the presence and absence of the bands (*GSTT1*-null and *GSTM1*-null genotype, respectively), whereas *GSTP1* rs1695 and rs1138272 polymorphisms were assessed using codominant model. Additionally, logistic regression analyses were conducted to compute the potential confounder's influence of both genetic and environmental factors for gastric cancer. Then, the independent effect of risk factors was investigated in a multivariate model (introducing all variables and terms of interactions) keeping only those statistically significant or showing a confounding effect on the studied factors. The likelihood ratio test was used to decide whether to retain each covariate in the model. Sex, gender, cigarette smoking, family history of gastric cancer, tuibur consumption, alcohol intake, smoked meat/vegetable intake, high consumption of salt and pickle, high consumption of fermented fat, *H. pylori* status and genotyping, *EBV* status, and genotyping were all considered in the regression model as potential confounders to evaluate the association of risk factors and susceptibility to gastric cancer. For all tests, a two-sided p-value <.05 was considered statistically significant. All statistical analyses were performed using SPSS 20.0 program (SPSS Ibérica, Madrid, Spain) and SYSTAT 13.0. (Systat Software Inc., USA). Heat map analysis was performed by Hemi IBP software.

Multifactor Dimensionality Reduction (MDR) Analysis

MDR analysis (MDR 3.0.2) was performed to assess the risk of gastric cancer by studying the gene–gene and gene–environment effect on the basis of GST genotyping and *GSTP1* gene expression. It is a nonparametric, model-free statistical approach which is used to generate one-dimensional model to predict gastric cancer susceptibility. The test is used to overcome the limitations encountered by parametric methods like logistic regression (LR) due to small sample size. Data were generated using 10-fold cross-validation procedure and 10 times random seed number to reduce the chance of false positives. The best model was selected on the basis of maximum cross-validation consistency (CVC) and testing balance accuracy (TBA). All MDR results with $p \leq .05$ were considered to be significant [26].

Interaction Entropy Graph

Interaction entropy graph was constructed using MDR software package (MDR 3.0.2). The graphs comprised of a node for each variable with pairwise linkage between them. The percentage of entropy removed by each variable is visualized for each node. The entropy estimates calculated were used to interpret the independent effect of each variable in comparison with the interaction effect of the variables and to determine if the interactions were synergistic [27]. The potential confounding factors, used in the multivariate model, were considered for interaction entropy estimation.

Homology Modeling of GSTP1 Ile105Val (rs1695) and GSTP1 Val114Ala (rs1138272)

Healthy control and mutant GSTP1 protein sequences were submitted to the SWISS MODEL (http://swissmodel. expasy.org/) and Phyre2 server for the prediction of three-dimensional structure [28–30]. Phyre2 uses the hidden Markov method to generate alignments of a submitted protein sequence against proteins with published structures. The resulting alignments were used to predict homology-based three-dimensional structure. The model was judged to be accurate when over 90% of the submitted residues were modeled at greater than 90% confidence.

Results

Clinical and Demographic Characteristics

Gastric cancer was more prevalent in males (70%) among the Mizo-Mongloid population (Table 2). Infection with *H. pylori* and *CagA*+ strains was more frequent in patients with GC than in controls (OR: 1.168; 95% CI: 0.365–3.733; $p = .094$ and OR: 8.298; 95% CI: 10.365–6.578; $p < .005$, respectively).

Table 2 Demographic and clinical characteristics of the samples

Demographic factor	[a]HC (n = 80)	[b]GC (n = 80)	[c]ORs (95% CI)[d]	*p* value
Age years ± SD (range)	52.18 ± 12.35	58.7 ± 9.76	–	
Sex				
Male	40 (50%)	56 (70%)	–	
Female	40 (50%)	24 (30%)		
Cigarette smoking	52 (65%)	56 (70%)	0.595 (0.373–1.947)	0.029
H. pylori positive	32 (40%)	56 (70%)	1.168 (0.365–3.733)	0.094
CagA positive	8 (10%)	36 (45%)	8.298 (0.365–6.578)	0.005
EBV positive	6 (7.5%)	18 (22.5%)	1.332 (0.255–6.957)	0.084
Family history of gastric cancer	10 (12.5%)	12 (15%)	0.708 (0.674–2.030)	0.477
High salt and pickle intake	56 (70%)	62 (77.5%)	0.687 (0.244–1.936)	0.178
Tuibur consumption	28 (35%)	34 (42.5%)	0.828 (0.459–1.492)	0.049
Smoked meat/vegetable Intake	50 (62.5%)	72 (90%)	6.181 (2.528–12.346)	0.001
High fermented fat intake	28 (35%)	72 (90%)	3.387 (1.902–6.031)	0.0003
Alcohol intake	34 (42.5)	40 (50%)	0.543 (0.312–1.944)	0.031

[a]HC – Healthy Control.
[b]GC – Gastric Cancer.
[c]OR – Odds ratio.
[d]95% CI – 95% Confidence Interval.

However, no differences in the prevalence of *Epstein–Barr virus* (*EBV*) strains were observed between patients and controls (22.5 vs 7.5%, respectively). Among the various demographic factors in the bivariate analysis, extra intake of fermented fat (OR: 3.387; 95% CI: 1.902–6.031; $p < .0003$) and smoked meat consumption (OR: 6.181; 95% CI: 2.528–12.346; $p < .001$) was also found to be risk factors for GC (Table 2). No significance was found with consumption of alcohol and familial history with gastric cancer risk.

Genotyping

Significant differences in genotype distribution were observed between patients and control samples (Table 3). The frequencies of *GSTM1* (67.5 vs 35%) and *GSTT1* (47.5 vs 12.5%) null genotypes in patients with GC were significantly high from those observed in controls (Fig. 1A). Moreover, the simultaneous presence of both the *GSTM1* and *GSTT1* null genotypes was significantly different between patients and control samples (30% in GC vs 7.5% in HC; OR: 2.29, 95% CI: 1.36 – 9.53). Significant differences in genotype or allele frequencies of the *GSTP1* IIe105Val and *GSTP1* Ala114Val polymorphisms were also observed. Presence of heterozygous allele A/G in rs1695 was 32.5% in GC and 20% in HC (OR: 1.93; 95% CI: 0.70–5.33) and homozygous allele G/G in rs1995 was 27.5% in GC and 5% in HC (OR: 7.21; 95% CI: 1.48–13.07). In case of rs1138272, C/T heterozygous was 25% in GC and 15% in HC (OR: 1.88; 95% CI: 0.62–5.81) and T/T homozygous allele was 5% in GC and absent in HC (Table 3). We also examined whether plausible contribution of GST polymorphisms to GC risk could be modified by interaction with other risk factors (smoking habit, *H. Pylori* and *EBV* infection, and family history of GC) identified in our study population (Tables 3 and 4). The exposure variables were as follows: smoking habit (current smokers vs nonsmokers), *H. pylori* infection (positive vs negative), *CagA* amplification status (positive vs negative), and family history of GC (positive vs negative; positive history was defined as any reported GC in first-degree relatives or two or more GC cases in second-degree relatives). Smoking status or family history of GC did not modify the association between GST gene polymorphisms and GC risk (Table 2). *H. pylori* and *CagA* positive status was significantly associated with the GST polymorphism. Moreover, tests for interaction under a multiplicative model showed strong statistically significant interaction between GST genotypes and the risk factors.

In summary, logistic regression analysis identified *H. pylori* infection (OR: 1.168; 95% CI: 0.365–3.733; $p = .094$) with *CagA*+ strains (OR: 8.298; 95% CI: 0.365 –36.578; $p < .005$), smoking habit (OR: 0.595; 95% CI: 0.373–1.947; $p = .029$) as independent risk factors for the development of GC (Fig. 1B, Table 2). Concerning the genetic factors, *GSTT1* and *GSTM1/T1* null genotype is strongly related with the GC phenotype with OR:

Table 3 Genotype pattern for GST gene polymorphism in the samples

Genotype	Total HC[a] (n = 80)	Total GC[b] patients (n = 80)	[c]OR	[d]95% CI
GSTM1				
(+)	52 (65%)	26 (32.5%)		
(−)	28 (35%)	54 (67.5%)	0.208	1.53–9.75
p value = .973				
GSTT1				
(+)	70 (87.5%)	42 (52.5%)		
(−)	10 (12.5%)	38 (47.5%)	0.389	0.052– 2.924
p value = 1.05				
GSTM1/T1				
(+/+)	44 (55%)	12 (15%)	0.043	0.008– 0.249
(−/+)	6 (7.5%)	14 (17.5%)	0.389	0.052– 2.924
(+/−)	24 (30%)	30 (37.5%)	0.208	0.039– 1.116
(−/−)	6 (7.5%)	24 (30%)	2.29	1.36–9.53
p value = .049				
GSTP1 105 (rs1695)				
A/A	60 (75%)	28 (35%)	0.22	0.09–0.58
A/G	16 (20%)	32 (32.5%)	1.93	0.70–5.33
G/G	4 (5%)	20 (25%)	7.21	1.48–13.07
p value = .057				
GSTP1 114 (rs1138272)				
C/C	68 (85%)	44 (55%)	0.22	0.74–0.63
C/T	12 (15%)	20 (25%)	1.88	0.62–5.81
T/T	0 (0%)	16 (5%)	NA	NA
p value = .261				

[a]HC – Healthy Control.
[b]GC – Gastric Cancer.
[c]OR – Odds ratio.
[d]95% CI – 95% Confidence Interval.
p value (<.05) indicates significant association of the genotype with gastric cancer.

0.389; 95% CI: 0.052–2.924, and OR: 2.29; 95% CI: 1.36–9.53, respectively. *GSTP1*105 (rs1695) was also significantly related to GC phenotype (Table 3). Significant association of GST polymorphisms and *H. pylori* infection with *CagA* was observed. Presence of *GSTT1* null alleles in patients with GC was significantly associated with the presence of *H. pylori* and *CagA* (OR: 3.00; 95% CI: 0.85–10.54 and OR: 6.65; 99% CI: 1.98–2.27). Presence of *GSTM1* null alleles in patients with GC was also significantly associated with presence of *CagA* genotype (OR: 5.93 95% CI: 1.53–2.90). *GSTM1/T1* null genotyping was significantly associated with the presence of *H. pylori* (OR: 4.67 95% CI: 1.19–8.33) and *CagA* (OR: 3.08 95% CI: 0.75–2.61) (Fig. 1B, Table 4). Multiplex PCR analysis of *GSTM1/T1* and *GSTP1* polymorphisms gave 0.767 area under the curve value (AUC) with a sensitivity of 59% and specificity of 76%, respectively (Fig. 2A). In addition, validation of *GST* polymorphisms associated with food habits (AUC = 0.721, sensitivity- 85%, specificity - 60%) and *H. pylori* (AUC = 0.867, sensitivity - 90%, specificity - 65%) was found to be significant (Fig. 2B, 2C).

MDR Analysis

MDR analysis was used to further validate our logistic regression based study on gene–gene and gene–environment interaction using a model-free approach. A four-order interaction model was chosen along with their cross-validation consistency (CVC) and testing balance accuracy (TBA). The best model was selected based on the highest CVC and TBA value among all the models. Among the entire dataset, the presence of *H. pylori* with smoked meat consumption is the best one-factor model that was found to be statistically significant ($p < .001$) with a CVC of 10/10 and testing accuracy of 0.54. The combination of *GSTT1* null genotype and *H. pylori* presence was found to be the best two-factor model with CVC of 5/10 and TBA of 0.33 ($p < .001$). The combination of *GSTT1* null genotype, *H. pylori* with *CagA* genotype, and smoked meat consumption and smoking was found to the best three-factor model, which was also the best overall model with a CVC of 10/10 and TBA of 0.54 ($p < .001$). Presence of G/G genotype (−0.141) of *GSTP1* rs1695 and presence of T/T genotype (−0.183) of *GSTP1* rs1138272 were significantly associated with less expression of *GSTP1* gene (Fig. 3B).

Interaction Entropy Graph

The interaction entropy graph shows the presence of *CagA* genotyping of *H. pylori* (14.2%) with smoked meat or vegetable and oral snuff as the large independent effectors in gastric cancer among the various environmental factors (Fig. 3A). The *GSTT1* null genotype (7.7%) was found to contribute the highest independent effect among all the genetic factors, while a synergistic effect was found in the case of *GSTT1* gene with *H. pylori* and *CagA* genotyping by removing 0.15 and 2.06% of entropy. Consumption of smoked meat was also found to contribute small independent entropy of 1.66%. G/G and A/G genotype of *GSTP1* rs1695 showed high degree of synergistic interaction with the *GSTP1* gene expression (Fig. 3B).

Comparative Modeling of GSTP1 Polymorphisms

The natural variants were substituted into the wild-type sequence for comparative modeling. The variance of *GSTP1* Ile105Val (rs1695) and Val114Ala (rs1138272) indicates interruption of H-site substrate binding pocket

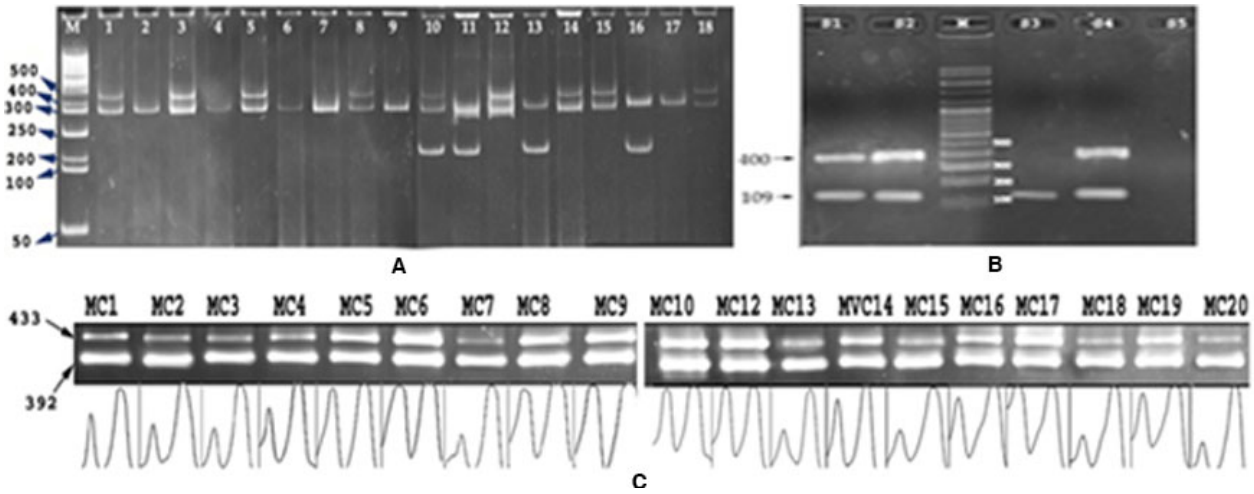


Figure 1 (*A*) Multiplex PCR amplified products of *GSTT1* (459 bp), *GSTM1* (219 bp), and internal control gene albumin (349 bp); (*B*) *16S rRNA* (109 bp) and *CagA* (400 bp) region of *H. pylori* amplified by multiplex PCR; (*C*) *GSTP1* (433 bp) expression in 10% polyacrylamide gel, β-Actin (392 bp) as control. M – Low range ruler plus (100–3000 bp); (*A*) 1–9 Gastric cancer Sample; 10–18 Healthy control Samples; (C - Digested fragment); S4 – AciI RFLP (T-Intact fragment); (*B*) S2 to S4 – *H. pylori* positive samples, S1 and S5 – positive and negative control, respectively; S3 – *H. pylori* positive but CagA negative.; (*C*) MC1 to MC20 – gastric cancer sample.

for chemical binding and was found to affect the metabolizing activity of the protein. Estimated protein folding by Phyre2 and Swiss model servers also differed from normal and mutated *GSTP1* (Fig. S1).

Densitometric Analysis

Densitometric analysis of *GSTP1* expression in relation to *GSTP1* genotyping for healthy control and gastric cancer individuals was performed by ImageJ and represented in Heat map by HEMI 1.0.1 software. Majority of the gastric cancer sample showed less *GSTP1* expression than healthy controls (Fig. 1C, 4A, B). Samples (25%) exhibiting low expression of *GSTP1* showed homozygous mutant type G/G and T/T for rs1695 and rs1138272, respectively. Heterozygous A/G and C/T for *GSTP1* rs1695 and rs1138272, respectively, had moderate expression. The A/A for *GSTP1* 105 (rs1695) and C/C for *GSTP1* 114 (rs1138272) wild-type genotyping showed approximately similar expression like healthy controls (Fig. 4A, B).

Discussion

In our study, *GSTM1/T1* gene deletions were found to be associated with a higher risk of GC. Moreover, the simultaneous presence of both the *GSTM1* and *GSTT1* null genotypes was not identical in GC and control groups (30% in GC vs 7.5% in HC; OR: 2.29, 95% CI: 1.36–9.53). A significant increase in risk of GC for individuals with deletions in both *GSTM1* and *GSTT1* genes was observed by collecting data from seven different planned studies (319 cases and 656 controls) that were *GSTT1* and *GSTM1* genotype combinations [31–33]. However, simultaneous absence of *GSTM1* and *GSTT1* genes was not significantly associated to GC risk in some other studies carried out in different geographic populations [34].

In the *GSTP1-1* gene, *GSTP1* 105 (rs1695) polymorphism evaluated was related to GC susceptibility or phenotype. The homozygous status of *GSTP1* 105 (rs1695) (A>105G) was most susceptible to GC risk in the Mizo population (OR: 7.21; 95% CI: 1.48–35.07). The allele frequencies of the *GSTP1* Ile105Val (rs1695) and Val114Ala (rs1138272) polymorphisms were not similar to European or Western populations [35], but similar in other Asian population [36]. The variant *GSTP1* Val105 allele was found to be linked to the risk of esophageal squamous cell carcinoma with a predisposition associated with tobacco smoking in Swedish patients [37]. A probabilistic description for this inconsistency is the important ethnic and geographic variation observed in the GST profile. Such variation is of particular interest in the case of *GSTM1* and *GSTT1* null frequencies, which differ considerably between Asians and Caucasians. These geographic variations in allele frequencies of GST gene along with reasonable connections with other gene polymorphisms could be the probable explanation for these observations.

A large number of epidemiological studies have reported the association between food habit, tobacco

Table 4 Mediation effect of *H. pylori* infection and their association with GST gene polymorphism

		H. pylori Negative		*H. pylori* Positive		*CagA* status Negative		*CagA* status Positive	
Polymorphism	Genotype	HC/GC	OR (95% CI)	HC/GC	OR (95% CI)	HC/GC	OR (95% CI)	HC/GC	OR (95% CI)
GSTM1	+	38/8	1 (Reference)	14/18	1.37 (0.45–4.12)	50/16	1 (Reference)	2/10	5.57 (0.62–0.03)
	–	10/16	0.179 (0.02–1.61)	18/38	3.11 (1.8–8.19)	22/28	0.33 (0.09–1.17)	6/26	5.93 (1.53–2.90)
			p value for interaction = .125				*p* value for interaction = .087		
GSTT1	+	46/6	1 (Reference)	24/36	1.90 (0.76–4.78)	70/40	1 (Reference)	0/2	NA
	–	2/18	0.871 (0.31–2.44)	8/20	3.00 (0.85 –10.54)	4/2	0.894 (0.35–2.26)	8/34	6.65 (1.98–2.27)
			p value for interaction = .793				*p* value for interaction = .813		
GSTM1/T1	+/+	28/8	1 (Reference)	16/4	0.21 (0.04–1.16)	44/8	1 (Reference)	0/4	NA
	–/+	2/6	3.16 (0.31–3.77)	4/8	2.11 (0.36–2.45)	6/4	0.64 (0.10–4.11)	0/10	NA
	+/–	18/6	0.27 (0.06–1.12)	6/24	5.28 (1.36–0.53)	22/24	1.12 (0.42–2.97)	2/6	3.16 (0.31–1.77)
	–/–	0/2	NA	6/22	4.67 (1.19–8.33)	0/8	NA	6/16	3.08 (0.75–2.61)
			p value for interaction = .046				*p* value for interaction = .053		
GSTP1 105rs1695	A/A	46/8	1 (Reference)	14/24	2.02 (0.70–5.82)	58/24	1 (Reference)	2/8	4.33 (0.46–0.60)
	Carrier G	2/16	9.75 (1.15–8.10)	18/32	2.29 (0.87–6.08)	14/20	1.57 (0.53–4.65)	6/28	6.64 (1.73–5.46)
			p value for interaction = .042				*p* value for interaction = .035		
GSTP1 114rs1138272	C/C	48/10	1 (Reference)	20/34	2.21 (0.85–5.74)	64/18	1 (Reference)	4/26	9.14 (1.90–3.89)
	Carrier T	0/14	NA	12/22	2.14 (0.70–6.53)	8/26	4.33 (1.27–4.77)	4/10	2.71 (0.49–4.90)
			p value for interaction = .485				*p* value for interaction = .582		

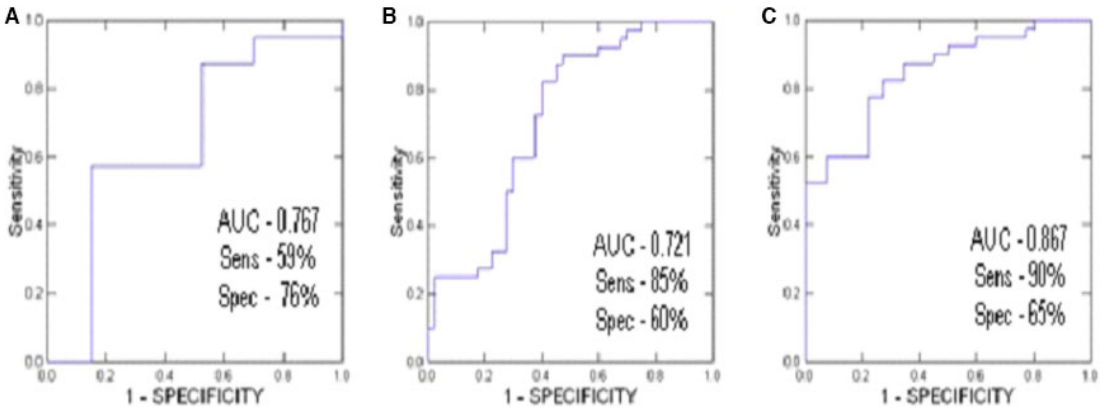


Figure 2 Receiver operating characteristic curve based on (*A*) multiplex analyses of GSTM1, GSTT1, and GSTP1 polymorphisms; (*B*) GST polymorphism associated with food habits; (*C*) GST polymorphism associated with *H. pylori* The predicted probability from the final logistic regression model is calculated to differentiate between healthy control and Mizoram gastric cancer. AUC – area under the curve; Sens – Sensitivity; Spec – Specificity.

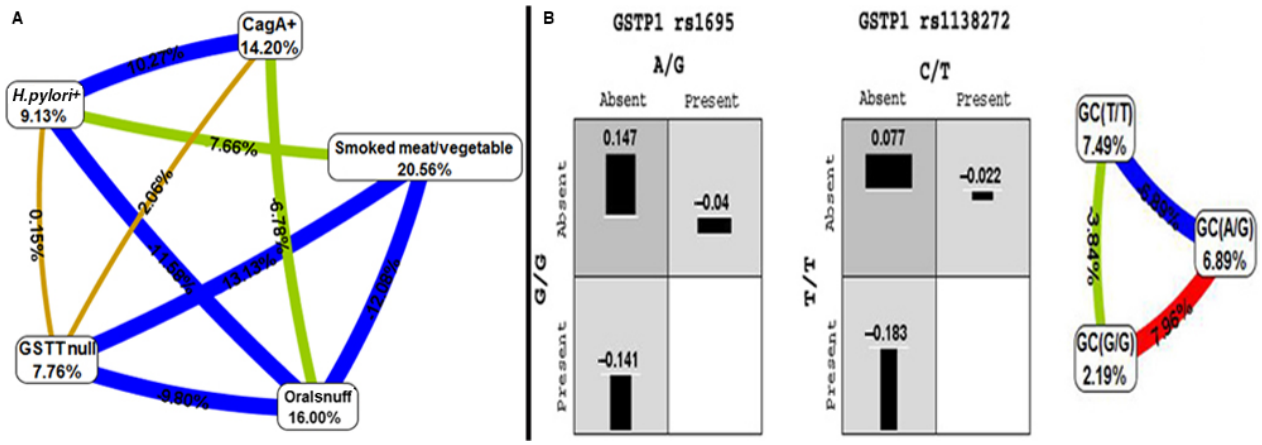


Figure 3 (*A*) Interaction entropy graph to find the gene–environment interaction with gastric cancer risk. The percentage of the entropy for independent factors as well as their interactions is represented in the graph. (*B*) Summary of the two SNPs GSTP1 rs1695: A>G (A/G and G/G) and GSTP1 rs1138272: C>T (C/T and T/T) relation with GSTP1 gene expression in MDR analysis. Black bar height shows GSTP1 expression. Bar wide shows frequencies of sample. Positive percentage of entropy denotes synergistic interaction, while negative percentage denotes redundancy. Here, the red color denotes the high degree of synergistic interaction, golden color denotes the moderate degree of synergistic interaction, green color represents moderate redundancy, while blue color denotes the highest. GC – gastric cancer.

smoking, and GC risk [38–42]. However, very few studies have published the potential interaction between tobacco smoking and GST gene polymorphisms with respect to GC risk [43,44]. In the present study, we aimed to investigate whether the reasonable GST gene variants associated with GC risk could be triggered by other environmental risk factors like *H. pylori* infection, consumption of smoked food, and tobacco related products. In our study, smoking habit was not significantly associated between *GST* variants and GC risk when compared with other factors. This finding is in concurrence to a recent meta-analysis where no association between the *GSTT1* and *GSTM1* null genotype with increased risk of GC in either smokers or nonsmokers was observed [45,46].

In accordance to previous reports, *H. pylori* infection was identified as a potential risk factor associated with GST polymorphisms and GC risk [47–49]. A significant effect of *H. pylori* status was observed in modifying the contribution of GST polymorphisms to GC risk. Palli et al. reported a significant increase of GC among *H. pylori*-infected individuals with deletions in both the *GSTM1* and the *GSTT1* genes in a high-GC risk Italian population [50]. In the present study, *CagA*+ strain was highly associated with *GSTM1/T1* null allele and *GSTP1* 105 (rs1695) with OR: 3.08; 95% CI 0.75–2.61 and OR:

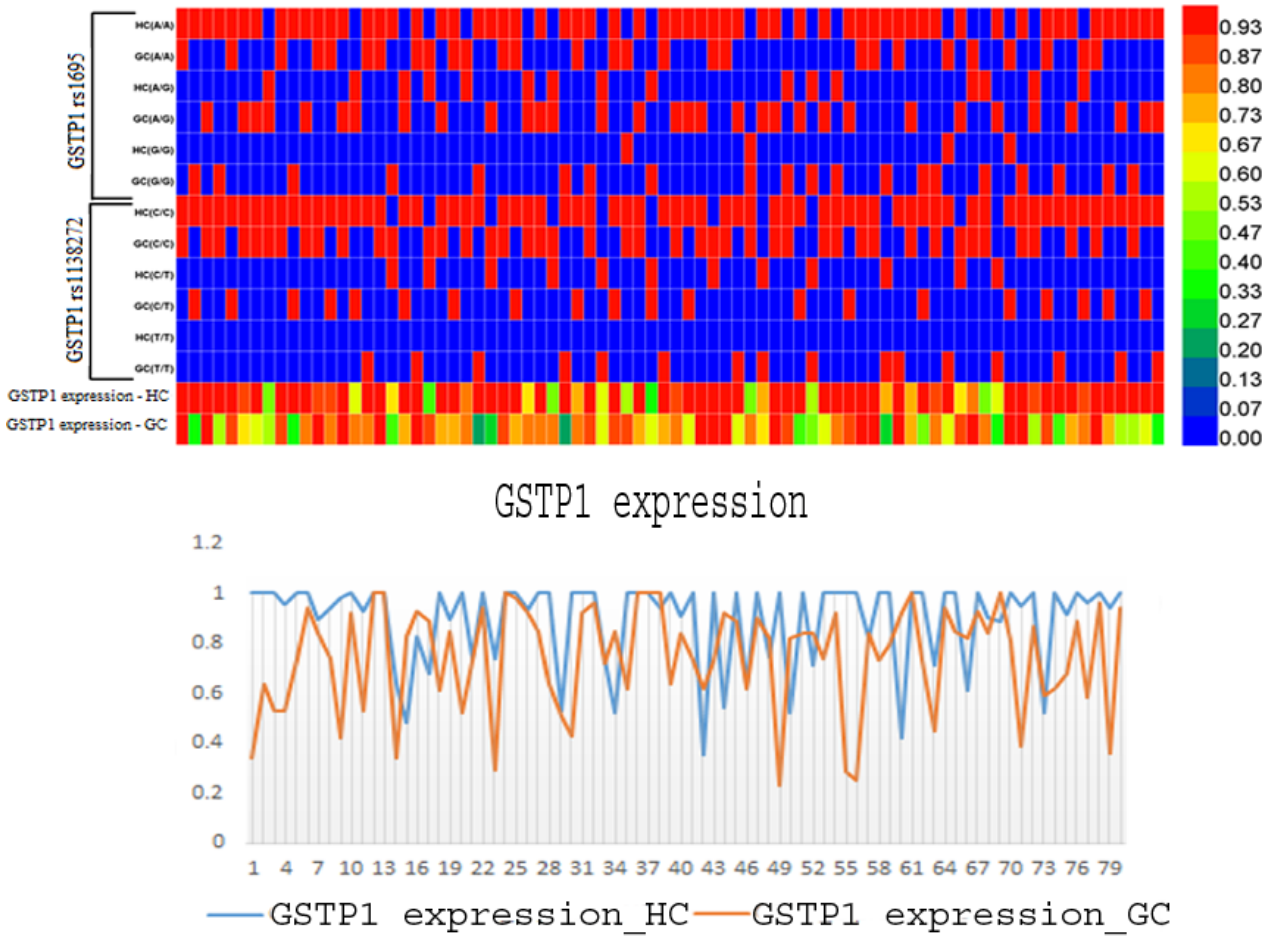


Figure 4 Heat map and graphical representation of GSTP1 gene expression relation with two different SNPs and their genotyping. HC – healthy control, GC – gastric cancer.

6.64 95% CI: 1.73–5.46, respectively (Table 4). In addition, studies on *GSTP1* allelotypes found that the isoenzymes with valine in position 105 (*GSTP1*) were able to metabolize the diol epoxides of polycyclic aromatic hydrocarbons more efficiently and those with a A114V transition in the presence of V105 are able to more effectively metabolize benzo(a)pyrene (Fig. S1) [51,52]. The MDR analyses also explained the gene–environment interaction (Fig. 3A) and identified the combination of presence of *H. pylori* with *CagA* genotype and smoked meat consumption as the best model for gastric cancer risk associated with *GSTT1* null genotype in Mizo population (Fig. 3A, Table 2). Interaction entropy graph was drawn for visualization and interpretation of MDR interactions. Presence of *H. pylori* with *CagA* genotype showed significant individual effect with *GSTT1* gene polymorphisms as well as strong synergistic effect among each other in gastric cancer risk and validated the results of gene–environment interaction. The relatively small sample size in our study might be a shortcoming for predicting high-order interactions and multiple comparison analysis; however, MDR approach improves statistical power to overcome small sample size limitation using cross-validation and permutation testing strategy. *GSTP1*, on the other hand, is expressed at high levels in extra hepatic tissues, including the kidney, liver, and the gastrointestinal tract suggesting an important role in the protection against carcinogens and other xenobiotics in these tissues [13]. Abnormal or low expression might not protect the cells from carcinogens or other xenobiotics [13]. In this study, majority of the samples showed significant negative *GSTP1* expression for gastric cancer samples in the presence of G/G, A/G (rs1695), and T/T (rs1138272) genotypes. The mutant homozygous condition for these two rs-IDs might be potential risk factor of low *GSTP1* expression followed by gastric cancer development in Mizo population. Low level *GSTP1* expression had significant

relation with homozygous mutant type G/G for GSTP1 105 (rs1695) and T/T for GSTP1 114 (rs1138272) (Fig. 4A, B). These results indicate that an individual's detoxification capacity and exonic mutation in the xenobiotic gene also play a crucial role in the initiation of gastric cancer [53]. However, no differences in the occurrence of *Epstein–Barr virus* (*EBV*) strains were observed between patients and controls associated with the *GSTP1* gene polymorphisms. After adjustment for potential confounding factors, a statistically significant difference was observed in the frequency of GSTM1/T1 null genotyping with the presence of *H. pylori* and CagA like other study [54]. Extra intake of fermented fat ($p < .0003$) and smoked meat consumption ($p < .001$) was also found to be potential confounders for gastric cancer risk in relation to GSTM1/T1 null genotyping and GSTT1 null genotyping. *GSTP1*105 (rs1695) for homozygous G/G was also significantly related to GC phenotype as a potential confounder like other population [55]. Further adjustment for other potential confounders did not substantially change the estimated ORs; therefore, we present only the estimates adjusted for significant matching variables.

In this study, the limitations concerning sample size can be ignored because of the fact that the Mizo ethnic group has a very small population size of about 10, 91,014 with 5,52,339 males and 5,38,675 females [56]. The fact that our study is one of the first performed in the unique and high gastric cancer risk Mizo-Mongoloid population from northeast India will provide insights for early diagnosis of this critical disease as well help to evaluate the genetic epidemiology of the disease for other populations as well.

In conclusion, our data showed that *GSTM1*, *GSTT1*, and *GSTP1* polymorphisms analyzed in this study were found to be significantly associated with the demographic factors and GC risk in Mizo population. Similar to several other complex diseases, it is very devious to consider the weight of each demographic factor concerned in its pathogenesis separately, particularly the contribution of genetic factors. Presence of *H. pylori* with *CagA* genotype showed significant individual effect with *GSTT1* polymorphisms as well as strong synergistic effect among each other in gastric cancer risk and validated the results of gene–environment interaction. Significant negative expression of *GSTP1* for gastric cancer samples with presence of G/G, A/G (rs1695), and T/T genotype (rs1138272) might not protect the cell from carcinogens or other xenobiotics. Further, prospective studies with larger sample size analyzing gene–environment relations in different geographic areas, and ethnic groups are needed to assess the significance of each factor for gastric cancer.

Acknowledgements and Disclosures

This work was supported by the DBT-Twinning Project on Gastric Cancer (BT/360/NE/TBP/2012) and Bioinformatics Infrastructure Facility (BT/BI/12/060/2012 (NERBIF-MUA) sponsored by the Department of Biotechnology (DBT), New Delhi, Govt. of India. The authors are thankful to Genesis Laboratory, Aizawl, Mizoram for providing the clinical information.

Competing interests: The authors declare that they have no conflict of interest.

References

1 Murray CJ, Lopez AD. Mortality by cause for eight regions of the world: Global Burden of Disease Study. *Lancet* 1997;349:1269–76.
2 Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM. Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. *Int J Cancer* 2010;127:2893–917.
3 Parkin DM, Muir CS, Whelan SL, Ferlay J, Teppo L, Thomas DB. Cancer incidence in five continents Vol VIII. *IARC, Sci Pub* 2002;7:1555.
4 Correa P. Human gastric carcinogenesis: a multistep and multifactorial process-first American cancer society award lecture on cancer epidemiology and prevention. *Cancer Res* 2002;52:6735–40.
5 Phukan RK, Narain K, Zomawia E, Hazarika NC, Mahanta J. Dietary habits and stomach cancer in Mizoram, India. *J Gastroenterol* 2006;41:418–24.
6 Hayes JD, Strange RC. Glutathione S-transferase polymorphisms and their biological consequences. *Pharmacology* 2000;61:154–66.
7 Hayes JD, Pulford DJ. The glutathione S-transferase supergene family: regulation of GST and the contribution of the isoenzymes to cancer chemoprotection and drug resistance. *Crit Rev Biochem Mol Biol* 1995;30:445–600.
8 Lan Q, Chow WH, Lissowska J, Hein DW, Buetow K, Engel LS, Ji B, Zatonski W, Rothman N. Glutathione S-transferase genotypes and stomach cancer in a population-based case-control study in Warsaw, Poland. *Pharmacogenetics* 2001;11:655–61.
9 Pemble S, Schroeder KR, Spencer SR, Meyer DJ, Hallier E, Bolt HM, Ketterer B, Taylor JB. Human glutathione S-transferase theta (GSTT1): cDNA cloning and the characterization of a genetic polymorphism. *Biochem J* 1994;300:271–6.
10 Cai L, Yu SZ, Zhang ZF. Glutathione S-transferases M1, T1 genotypes and the risk of gastric cancer: a case-control study. *World J Gastroenterol* 2001;7:506–9.
11 Zimniak P, Nanduri B, Pikuła S, Bandorowicz-Pikuła J, Singhal SS, Srivastava SK, Awasthi S, Awasthi YC. Naturally occurring human glutathione S-transferase GSTP1-1 isoforms with isoleucine and valine in position 104 differ in enzymic properties. *Eur J Biochem* 1994;224:893–9.
12 Tahara T, Shibata T, Nakamura M, Okubo M, Yamashita H, Yoshioka D, Yonemura J, Ishizuka T, Hirata I, Arisawa T. Polymorphisms of DNA Repair and Xenobiotic Genes Predispose to CpG Island Methylation in Non-Neoplastic Gastric Mucosa. *Helicobacter* 2011;16:99–106.
13 DeBruin W, Wagenmans M, Peters W. Expression of glutathione S-transferase alpha, P1-1 and T1-1 in the human gastrointestinal tract. *Jpn J Cancer Res* 2000;91:310–6.

14 Garte S, Gaspari L, Alexandrie AK, et al. Metabolic gene polymorphism frequencies in control populations. *Cancer Epidemiol Biomarkers Prev* 2001;10:1239–48.
15 Mahanta J, Chetia M, Hazarika NC, Narain K, Sharma SK. Toxicity of tuibur, a unique form of tobacco smoke extract used in Mizoram India. *Curr Sci* 1998;75:381–4.
16 Russo A, Maconi G, Spinelli P, et al. Effect of lifestyle, smoking, and diet on development of intestinal metaplasia in *H. pylori*-positive subjects. *Am J Gastroenterol* 2001;96:402–8.
17 National Cancer Registry Programme (NCRP). Three year report of the population based cancer registries 2006-2008. (First report of 20 PBCRs in India), Bangalore, Indian Council. *Med Res*. 2010;11:71–8.
18 Ghatak S, Doris L, Mawia L, Sapkota R. Zothanpuia, Pautu JL, Muthukumaran RB, Senthil-Kumar N. Mitochondrial D-Loop and Cytochrome Oxidase Subunit I Polymorphisms among the Breast Cancer Patients of Mizoram, Northeast India. *Curr Genet* 2014a;60:201–12.
19 Ghatak S, Muthukumaran RB, Senthil-Kumar N. A simple method of genomic DNA extraction from human samples for PCR-RFLP analysis. *J Biomol Tech* 2013;24:224–31.
20 Ghatak S. Zothansanga, Pautu JL, Senthil-Kumar N. Co-Extraction and PCR based analysis of nucleic acids from Formalin-Fixed Paraffin-Embedded Specimens. *J Clin Lab Anal* 2015;29:485–92.
21 Arand M, Mühlbauer R, Hengstler J, Jäger E, Fuchs J, Winkler L. Oesch F.A multiplex polymerase chain reaction protocol for the simultaneous analysis of the glutathione S-transferase GSTM1 and GSTT1 polymorphisms. *Anal Biochem* 1996;236:184–6.
22 Ali-Osman F, Akande O, Antoun G, Mao JX, Buolamwini J. Molecular cloning, characterization, and expression in *Escherichia coli* of full-length cDNAs of three human glutathione S-transferase Pi gene variants. Evidence for differential catalytic activity of the encoded proteins. *J BiolChem* 1997;272:10004–12.
23 García-G MA, Quintero E, Bujanda L, et al. Relevance of GSTM1, GSTT1, and GSTP1 gene polymorphisms to gastric cancer susceptibility and phenotype. *Mutagenesis* 2012;27:771–7.
24 Ho SA, Hoyle JA, Lewis FA, Secker AD, Cross D, Mapstone NP, Dixon MF, Wyatt JI, Tompkins DS, Taylor GR. Direct polymerase chain reaction test for detection of *Helicobacter pylori* in humans and animals. *J ClinMicrobiol* 1991;29:2543–9.
25 Kingma DW, Weiss WB, Jaffe ES, Chadburn A, Knowles DM. Epstein-Barr virus latent membrane protein-1 oncogene deletions: Correlations with malignancy in Epstein-Barr virus-associated lymphoproliferative disorders and malignant lymphomas. *Am J Pathol* 1997;151:805–12.
26 Hahn LW, Ritchie MD, Moore JH. Multifactor dimensionality reduction software for detecting gene-gene and gene-environment interactions. *Bioinformatics* 2003;19:376–82.
27 Manuguerra M, Matullo G, Veglia F, et al. Multi-factor dimensionality reduction applied to a large prospective investigation on gene-gene and gene-environment interactions. *Carcinogenesis* 2007;28:414–22.
28 Arnold K, Bordoli L, Kopp J, Schwede T. The SWISS-MODEL workspace: a web-based environment for protein structure homology modelling. *Bioinformatics* 2006;26:195–201.
29 Kiefer F, Arnold K. Ku¨nzli M, Bordoli L, Schwede T. The SWISS-MODEL repository and associated resources. *Nucleic Acids Res* 2009;37:387–92.
30 Kelley LA, Sternberg MJ. Protein structure prediction on the web: a case study using the Phyre server. *Nat Protoc* 2009;4:363–71.
31 Saadat M. Genetic polymorphisms of glutathione S-transferase T1 (GSTT1) and susceptibility to gastric cancer: a meta-analysis. *Cancer Sci* 2006;97:505–9.
32 Boccia S, La-Torre G, Gianfagna F, Mannocci A, Ricciardi G. Glutathione S-transferase T1 status and gastric cancer risk: a meta-analysis of the literature. *Mutagenesis* 2006;21:115–23.
33 Jing C, Huang ZJ, Duan YQ, Wang PH, Zhang R, Luo KS, Xiao XR. Glulathione-S-transferases gene polymorphism in prediction of gastric cancer risk by smoking and *Helicobacter pylori* infection status. *Asian Pac J Cancer Prev* 2012;13:3325–8.
34 Piao JM, Shin MH, Kweon SS, et al. Glutathione-S-transferase (GSTM1, GSTT1) and the risk of gastrointestinal cancer in a Korean population. *World J Gastroenterol* 2009;15:5716–21.
35 Zendehdel K, Bahmanyar S, McCarthy S, Nyren O, Andersson B, Ye W. Genetic polymorphisms of glutathione S-transferase genes GSTP1, GSTM1, and GSTT1 and risk of esophageal and gastric cardia cancers. *Cancer Cause Control* 2009;20:2031–8.
36 Mocellin S, Verdi D, Pooley KA, Nitti D. Genetic variation and gastric cancer risk: a field synopsis and meta-analysis. *Gut* 2015;64:1209–19.
37 Gonzalez CA, Pera G, Agudo A, et al. Smoking and the risk of gastric cancer in the European Prospective Investigation Into Cancer and Nutrition (EPIC). *Int J Cancer* 2003;107:629–34.
38 La-Torre G, Chiaradia G, Gianfagna F, De-Lauretis A, Boccia S, Mannocci A, Ricciardi W. Smoking status and gastric cancer risk: an updated meta-analysis of case-control studies published in the past ten years. *Tumori* 2009;95:13–22.
39 Ladeiras-Lopes R, Pereira AK, Nogueira A, Pinheiro-Torres T, Pinto I, Santos-Pereira R, Lunet N. Smoking and gastric cancer: systematic review and meta-analysis of cohort studies. *Cancer Cause Control* 2008;19:689–701.
40 Agudo A, Sala N, Pera G, et al. Polymorphisms in metabolic genes related to tobacco smoke and the risk of gastric cancer in the European prospective investigation into cancer and nutrition. *Cancer Epidemiol Biomarkers Prev* 2006;15:2427–34.
41 Soya SS, Vinod T, Reddy KS, Gopalakrishnan S, Adithan C. Genetic polymorphisms of glutathione-S-transferase genes (GSTM1, GSTT1 and GSTP1) and upper aerodigestive tract cancer risk among smokers, tobacco chewers and alcoholics in a an Indian population. *Eur J Cancer* 2007;43:2698–706.
42 Malakar M, Devi KR, Phukan RK, Kaur T, Deka M, Puia L, Barua D, Mahanta J, Narain K. Genetic polymorphism of glutathione S-transferases M1 and T1, tobacco habits and risk of stomach cancer in Mizoram, India. *Asian Pac J Cancer Prev* 2012;13:4725–32.
43 Malik MA, Upadhyay R, Mittal RD, Showket AZ, Dinesh RM, Balraj M. Role of xenobiotic-metabolizing enzyme gene polymorphisms and interactions with environmental factors in susceptibility to gastric cancer in Kashmir Valley. *J Gastrointest Cancer* 2009;40:26–32.
44 Katoh T, Nagata N, Kuroda Y, Itoh H, Kawahara A, Kuroki N, Ookuma R, Bell DA. Glutathione S-transferase M1 (GSTM1) and T1 (GSTT1) genetic polymorphism and susceptibility to gastric and colorectal adenocarcinoma. *Carcinogenesis* 1996;17:1855–9.
45 Boccia S, Sayed-Tabatabaei FA, Persiani R, et al. Polymorphisms in metabolic genes, their combination and interaction with tobacco smoke and alcohol consumption and risk of gastric cancer: a case-control study in an Italian population. *BMC Cancer* 2007;7:206.

46 Zhao Y, Deng X, Song G, Qin S, Liu Z. The GSTM1 null genotype increased risk of gastric cancer: a meta-analysis based on 46 studies. *PLoS ONE* 2013;8:e81403.
47 Palli D, Masala G, Del-Giudice G, et al. CagA+ *Helicobacter pylori* infection and gastric cancer risk in the EPIC-EURGAST study. *Int J Cancer* 2007;120:859–67.
48 Parsonnet J, Friedman GD, Vandersteen DP, Chang Y, Vogelman JH, Orentreich N, Sibley RK. *Helicobacter pylori* infection and the risk of gastric carcinoma. *N Engl J Med* 1991;325:1127–31.
49 Uemura N, Okamoto S, Yamamoto S, Matsumura N, Yamaguchi S, Yamakido M, Taniyama K, Sasaki N, Schlemper RJ. *Helicobacter pylori* infection and the development of gastric cancer. *N Engl J Med* 2001;345:784–9.
50 Palli D, Saieva C, Gemma S, et al. GSTT1 and GSTM1 gene polymorphisms and gastric cancer in a high-risk Italian population. *Int J Cancer* 2005;115:284–9.
51 Hu X, Xia H, Srivastava SK, Herzog C, Awasthi YC, Ji X, Zimniak P, Singh SV. Activity of four allelic forms of glutathione S-transferase hGSTP1-1 for diol epoxides of polycyclic aromatic hydrocarbons. *Biochem Biophys Res Commun* 1997;238:397–402.
52 Eaton DL, Bammler TK. Concise review of the glutathione S-transferases and their significance to toxicology. *Toxicol Sci* 1999;49:156–64.
53 Roth MJ, Abnet CC, Johnson LL, Mark SD, Dong ZW, Taylor PR, Dawsey SM, Qiao YL. Polymorphic variation of CYP1A1 is associated with the risk of gastric cardia cancer: a prospective case–cohort study of cytochrome P-450 1A1 and GST enzymes. *Cancer Cause Control* 2004;15:1077–83.
54 Zhang Y, Sun LP, Xing CZ, Xu Q, He CY, Li P, Gong YH, Liu YP, Yuan Y. Interaction between GSTP1 Val allele and *H. pylori* infection, smoking and alcohol consumption and risk of gastric cancer among the Chinese population. *PLoS ONE* 2012;7:e47178.
55 Lao X, Peng Q, Lu Y, Li S, Qin X. Zhiping Chen Z, Chen J. Glutathione S-transferase gene GSTM1, gene-gene interaction, and gastric cancer susceptibility: evidence from an updated meta-analysis. *Cancer Cell Int* 2014;14:127.
56 Anon. Chapter 2: Size, Growth Rate and Rural-Urban Distribution of Populations in Mizoram, Provisional Populations Totals – Mizoram. India: Census of India, 2011;pp:7–13.

Supporting Information

Additional Supporting Information may be found in the online version of this article:

Figure S1. Protein 3D structure of GSTP1. rs1695 - (A) healthy control; (B) Mutant (I105>V); rs1138272 - (C) healthy control; (D) mutant (A114>V).

RESEARCH

Five year record on Cancer Incidence from a Diagnostic Centre in Mizoram, Northeast India

Mary Vanlalhruaii Tonsing[1], Souvik Ghatak[1], Freda Lalrohlui[1], Nachimuthu Senthil Kumar[1] and John Zohmingthanga[2]*

[1]Department of Biotechnology, Mizoram University, Aizawl -796004, Mizoram, India
[2]Office of the Medical Superintendent, Civil Hospital, Aizawl -796001, Mizoram, India

*Corresponding author email: johnzo05@yahoo.co.in, nskmzu@gmail.com.

ABSTRACT

Cancer has become leading cause of death in Northeast Indian population. The main reason being poor knowledge of prevention and diagnosis combined with modern lifestyle of the Mizo population. All cancers have been reported in Mizo population including the cancers of stomach, cervix, lungs, breast, oesophagus, rectum, prostate, liver, bladder, oral etc. The cause of such high incidence rates of these cancers may be inherited or genetic and environmental factors such as life style and food habits, especially high consumption of tobacco and alcohol. A peculiar habit of tobacco smoke-infused water (Tuibur) is also in practice in this population. In view of these facts, the present article describes the status of various types of cancers in Mizo population. Besides, attempts have been made to describe the main causes of cancer in this population with their frequency and grading. In this study, increasing number of cancer patients in different age group was observed from 2011 – 2015. The highest incidence of cancer was observed in the patients with age group 50 - 60, followed by age groups above 60 years. In age group 20-30 years, the breast and cervix cancers were more prevalent from 2011 – 2015. In middle age group (30-40 and 40-50 years), the cervix, stomach and oesophagus cancers were more prevalent. The common type of cancer in female includes cervical cancer, followed by breast cancer which is shown to be increasing with age of the patients and also increasing each year within this 5 year period from 2011 – 2015. In 2014 and 2015, adenocarcinoma (AC) and squamous cell carcinoma (SCC) are both commonly seen. This study will help in understanding the etiology of cancer and also in developing preventive measures in future.

KEYWORDS: Cancer Prevalence, life style habits, Adenocarcinoma, Squamous cell carcinoma, Mizo Population, Northeast India

Citation: Tonsing MV et al (2019) Five year record on cancer incidence from a diagnostic centre in Mizoram, Northeast India. Cancer Health Disparities 3:e1-e15. doi:10.9777/chd.2019.1003.

INTRODUCTION

Cancer is a major public health problem in India. Cancer is the second and third most common cause of death in the developed and developing countries, respectively [Bener et al.,2008]. A large network of population based cancer registries in India, provides valuable data for cancer. Population-based survival comparisons for cancer between nine Asian countries showed that India has the lowest 5-year survival for most cancer sites; 5-year survival for breast cancer was 52% and for colorectal cancer was 28%, compared with 82% and 44% in China. Breast cancer (145000 cases per year), tobacco-related head and neck cancers (141000), cervical cancer (123000), lung cancer (70000), large bowel cancer (64000), and stomach cancer (63000) account for more than half of the burden, implying that prevention, along with early detection and treatment, are important interventions for cancer control [Bener et al.,2008]. About 70% cancer cases have been diagnosed, with little survival of the patients, over the last decades. The poor prospects of cancer survival indicates inadequate health care financing, and also because cancer diagnosis and treatment are becoming increasingly unaffordable for health-care systems in India as well as in many low-income and middle-income countries since it has centred on expensive diagnostic and staging investigations such as imaging, and on specialised treatments and costly drugs. Most frequently observed cancers in Indian population are lungs, breast, colon, rectum, stomach and liver [NCRP, 2010; Rao, 1998; Murthy et al., 2004]. It is important to study the status of cancers in India so that advance measures may be taken to control this havoc in near future. In view of these facts, attempts have been made to study the status of cancers in India including its causes, preventive measures, effect on Indian economy and comparison with global scenario.

DATA SOURCES AND METHODS

Study area

The study area Mizoram is flanked by Bangladesh on the west and Myanmar on the east and south. The total area is 21,081 sq. km. It mainly consists of 8 districts, namely Aizawl, Lunglei, Champhai, Lawngtlai, Mamit, Kolasib, Serchhip and Saiha. The total population of the State is 10,97,206 with about 5,55,339 males and 5,41,867 females as per the 2011 census. Patients from other districts also visit the capital city, Aizawl for diagnosis and treatment.

Data collection

Genesis Laboratory, Aizawl acts as the major diagnostic facility of cancer. Ethical approval was obtained from the Institutional Ethics Committee. The study subjects included 1,477 cancer patients diagnosed/registered in Genesis Laboratory, Aizawl between 2011 – 2015. This data included different types of cancer and the type of cells being affected both in males and females, within different age groups and its staging on the basis of clinical symptoms and histopathology at the Genesis Laboratory, Aizawl, Mizoram. The patients coming from all the different districts of Mizoram were included in this analysis.

Statistical analysis

The frequencies of the collected data were represented according to their age group, type of cancer, staging and sex ratio. The association in each group was estimated using odds ratios (ORs) and 95% confidence intervals (CIs). The association of various cancers and stages of the patients was tested for Hardy–Weinberg equilibrium by a chi-square test with one degree of freedom (df).

Table 1. List of various cancers year wise, stratified by age group.

Age group	Cancer type	No (%)	Male	Female	ORs (95% CI)	P value
2011						
20-30	Breast[a]	3 (60)	0	3	2.00 (0.42 - 9.33)	0.0497
	Liver[b]	1 (20)	0	1	0.20 (0.03 – 1.29)	
	Colon[b]	1 (20)	0	1	0.20 (0.03 – 1.29)	
31-40	Breast[ab]	6(24)	0	6	0.31 (0.12 – 0.76)	0.0159
	Uterus[b]	1(4)	0	1	0.92 (0.42 – 1.98)	
	Cervix[a]	10(40)	0	10	0.13 (0.04 – 0.42)	
	Endometrium[b]	2(8)	0	2	0.08 (0.02 – 0.33)	
	Colon[b]	1(4)	0	1	0.04 (0.007 - 0.24)	
	Stomach[b]	1(4)	1	1	0.04 (0.007– 0.24)	
41-50	Cervix[a]	13(22.4)	0	13	0.08 (0.03 – 0.18)	0.0001
	Tongue[c]	1(1.72)	1	0	0.26 (0.15 – 0.45)	
	Stomach[a]	13(22.4)	6	7	0.12 (0.06 – 0.25)	
	Oesophagus[ab]	11(18.96)	10	1	0.01 (0.002– 0.06)	
	Pyriform fossa[bc]	3(5.17)	3	0	0.17 (0.09 – 0.32)	
	Breast[abc]	4(6.89)	0	4	0.03 (0.01 – 0.11)	
	Nasopharynx[bc]	3(5.17)	1	2	0.03 (0.01 – 0.11)	
	Liver[abc]	4(6.89)	3	1	0.01 (0.002– 0.06)	
	Lung[abc]	5(8.62)	3	2	0.01 (0.002– 0.06)	
51-60	Oesophagus[c]	1(2.56)	0	1	0.04 (0.01 – 0.12)	0.0016
	Pharynx[ab]	13(33.33)	2	11	0.20 (0.11 – 0.37)	
	Breast[abc]	7(17.94)	7	0	0.10 (0.04 – 0.21)	
	Uterus[a]	14(35.89)	13	1	0.22 (0.12 – 0.40)	
	Cervix[c]	1(2.56)	1	0	0.01 (0.002– 0.07)	
	Lung[c]	3(7.69)	2	1	0.04 (0.01 – 0.12)	
	Stomach[c]	3(7.69)	2	1	0.04 (0.01 – 0.12)	
	Nasopharynx[c]	1(2.56)	1	0	0.01 (0.002– 0.07)	
	Tonsil[c]	1(2.56)	1	0	0.01 (0.01 – 0.07)	
	Endometrium[c]	1(2.56)	1	0	0.01 (0.01 – 0.07)	
	Liver[c]	1(2.56)	0	1	0.01 (0.01 – 0.07)	
	Pyriform fossa[c]	1(2.56)	0	1	0.01 (0.01 – 0.07)	
> 60	Stomach[a]	26(29.88)	14	12	0.10 (0.03 – 0.30)	0.0715
	Oesophagus[ab]	3(3.44)	3	0	0.03 (0.005– 0.18)	
	Nasopharynx[ab]	3(3.44)	3	0	0.73 (0.37 – 1.45)	
	Tonsil[b]	1(1.14)	1	0	0.10 (0.03 – 0.30)	
	Breast[ab]	3(3.44)	0	3	0.06 (0.01 – 0.24)	
	Cervix[ab]	3(3.44)	0	3	0.03 (0.005– 0.18)	
	Pyriform fossa[b]	1(1.14)	1	0	0.10 (0.03 – 0.30)	
	Tongue[b]	1(1.14)	1	0	0.03 (0.005– 0.18)	
	Urinary Bladder[b]	1(1.14)	1	0	0.03 (0.005– 0.18)	
	Epiglottis[b]	1(1.14)	1	0	0.03 (0.005– 0.18)	
	Lung[b]	26(29.88)	14	12	0.03 (0.005– 0.18)	
	Liver[a]	18(20.68)	11	7	0.03 (0.005– 0.18)	
2012						
20-30	Breast[a]	1(12.5)	0	1	0.05 (0.01 – 0.34)	0.0119
	Stomach[a]	2(25)	0	2	0.12 (0.03 – 0.48)	
	Colon[a]	1(12.5)	0	1	0.05 (0.01 – 0.34)	
	Cervix[a]	3(37.5)	0	3	0.05 (0.01 – 0.34)	
	Nasopharynx[a]	1(12.5)	0	1	1.25 (0.50 – 3.07)	
31-40	Liver[ab]	6(19.35)	2	4	0.09 (0.03 – 0.29)	0.0012

	Cervix[a]	7(22.58)	0	7	0.13 (0.04 – 0.36)	
	Breast[abc]	3(9.6)	0	3	0.25 (0.11 – 0.58)	
	Tongue[c]	1(3.22)	1	0	0.03 (0.005– 0.17)	
	Stomach[bc]	2(6.45)	1	1	0.03 (0.005– 0.17)	
	Oesophagus[a]	7(22.58)	6	1	0.06 (0.01 – 0.23)	
	Uterus[c]	1(3.22)	0	1	0.03 (0.005– 0.17)	
	Tonsil[c]	1(3.22)	1	0	0.06 (0.01 – 0.23)	
	Thyroid[c]	1(3.22)	0	1	0.03 (0.005– 0.17)	
	Pyriform fossa[c]	1(3.22)	1	0	0.41 (0.20 – 0.85)	
	Pharynx[c]	1(3.22)	1	0	0.03 (0.005– 0.17)	
41-50	Liver[a]	10(17.24)	5	5	0.18 (0.09 – 0.36)	0.0001
	Lung[abcd]	5(8.62)	5	0	0.18 (0.09 – 0.36)	
	Oesophagus[ab]	9(15.51)	8	1	0.20 (0.10 – 0.40)	
	Stomach[abc]	8(13.79)	8	0	0.01 (0.003– 0.10)	
	Nasopharynx[bcd]	3(5.17)	2	1	0.03 (0.009– 0.13)	
	Cervix[abcd]	7(12.06)	0	7	0.05 (0.01 – 0.16)	
	Vocal Cord[d]	1(1.72)	1	0	0.03 (0.009– 0.13)	
	Tonsil[d]	1(1.72)	1	0	0.01 (0.003– 0.10)	
	Larynx[d]	1(1.72)	1	0	0.01 (0.003– 0.10)	
	Uterus[d]	1(1.72)	0	1	0.01 (0.003– 0.10)	
	Pyriform fossa[d]	1(1.72)	1	0	0.09 (0.03 – 0.22)	
	Breast[d]	7(12.06)	0	7	0.11 (0.05 – 0.26)	
	Mouth[d]	1(1.72)	1	0	0.03 (0.009– 0.13)	
	Colon[cd]	2(3.44)	2	0	0.03 (0.009– 0.13)	
51-60	Lung[bc]	9(9.57)	5	4	0.10 (0.05 – 0.21)	0.0001
	Liver[c]	4(4.25)	2	2	0.47 (0.30 – 0.74)	
	Stomach[a]	29(30.85)	17	12	0.35 (0.21 – 0.57)	
	Tongue[c]	2(2.12)	2	0	0.05 (0.01 – 0.13)	
	Pyriform fossa[c]	4(4.25)	4	0	0.02 (0.006– 0.09)	
	Caecum[c]	2(2.12)	2	0	0.02 (0.006– 0.09)	
	Oesophagus[ab]	20(21.27)	18	2	0.01 (0.002– 0.06)	
	Nasopharynx[c]	1(1.06)	0	1	0.01 (0.002– 0.06)	
	Thyroid[c]	1(1.06)	0	1	0.02 (0.006– 0.09)	
	Pharynx[c]	2(2.12)	2	0	0.07 (0.03 – 0.17)	
	Tonsil[c]	3(3.19)	2	1	0.03 (0.01 – 0.11)	
	Cervix[c]	3(3.19)	0	3	0.01 (0.002– 0.06)	
	Uterus[c]	1(1.06)	1	0	0.02 (0.006– 0.09)	
	Colon[c]	1(1.06)	0	1	0.01 (0.002– 0.06)	
	Breast[bc]	7(7.44)	1	6	0.01 (0.002– 0.06)	
	Epiglottis[c]	3(3.19)	3	0	0.01 (0.002– 0.06)	
>60	Lung[a]	37(33.33)	15	22	0.04 (0.01 – 0.15)	0.0779
	Liver[b]	10(9.00)	5	5	1.08 (0.62 – 1.87)	
	Nasopharynx[b]	4(3.60)	3	1	0.25 (0.12 – 0.49)	
	Stomach[b]	36(32.43)	23	13	0.06 (0.02 – 0.19)	
	Cervix[b]	5(4.50)	0	5	0.04 (0.01 – 0.15)	
	Epiglottis[b]	2(1.80)	1	1	0.02 (0.003– 0.11)	
	Oesophagus[b]	10(9.00)	10	0	0.08 (0.03 – 0.23)	
	Breast[b]	3(2.70)	0	3	0.02 (0.003– 0.11)	
	Pyriform fossa[b]	2(1.80)	2	0	0.02 (0.003– 0.11)	
2013						
20-30	Stomach[a]	3(50)	0	3	0.20 (0.03 – 1.29)	0.1889
	Breast[a]	1(16.67)	0	1	1.00 (0.23 – 4.33)	
	Tongue[a]	2(33.33)	2	0	0.50 (0.10 – 2.33)	

31-40	Cervix[ab] Nasopharynx[c] Oesophagus[abc] Tongue[bc] Stomach[bc] Breast[a] Larynx[c] Tonsil[c] Uterus[bc]	12(21.42) 1(1.78) 8(14.28) 4(7.14) 5(8.92) 14(25.00) 1(1.78) 1(1.78) 4(7.14)	0 0 7 3 3 0 1 0 0	12 1 1 1 2 14 0 1 4	0.38 (0.21 – 0.71) 0.85 (0.49 – 1.47) 0.11 (0.45 – 0.27) 0.02 (0.003– 0.11) 0.02 (0.003– 0.11) 0.02 (0.003_ 0.11) 0.04 (0.01 – 0.15) 0.04 (0.01 – 0.15) 0.02 (0.003– 0.11)	0.0002
41-50	Lung[b] Liver[b] Cervix[ab] Stomach[ab] Caecum[b] Uterus[b] Breast[ab] Oesophagus[a] Pharynx[b] Bladder[b]	4(4.30) 5(5.37) 17(18.27) 14(15.05) 1(1.07) 2(2.15) 12(12.90) 27(29.03 1(1.07) 1(1.07)	0 2 0 10 0 0 0 23 1 0	4 3 17 4 1 2 12 4 0 1	0.31 (0.17 – 0.57) 0.61 (0.36 – 1.03) 0.20 (0.10 – 0.40) 0.01 (0.003– 0.10) 0.01 (0.003– 0.10) 0.05 (0.01 – 0.16) 0.01 (0.003– 0.10) 0.01 (0.003– 0.10) 0.07 (0.02 – 0.19) 0.01 (0.003– 0.10)	0.0001
51-60	Lung[b] Liver[b] Stomach[a] Epiglottis[b] Oesophagus[a] Tongue[b] Cervix[b] Skin[b] Gall Bladder[b] Tonsil[b] Pharynx[b] Breast[b] Pyriform fossa[b] Tongue[b] Colon[b] Nasopharynx[b] Bladder[b]	13(9.55) 10(7.35) 37(27.20) 2(1.47) 35(25.73) 1(0.73) 14(10.29) 1(0.73) 1(0.73) 1(0.73) 2(1.47) 6(4.41) 6(4.41) 1(0.73) 4(2.94) 2(1.47) 1(0.73)	7 5 17 2 30 1 0 1 0 1 2 0 6 1 2 2 1	6 5 20 0 5 0 14 0 1 0 0 6 0 0 2 0 0	0.06 (0.02 – 0.14) 0.25 (0.15 – 0.42) 0.54 (0.35 – 0.83) 0.06 (0.02 – 0.14) 0.04 (0.01 – 0.12) 0.02 (0.006– 0.08) 0.02 (0.006– 0.08) 0.02 (0.006– 0.08) 0.02 (0.006– 0.08) 0.01 (0.002– 0.06) 0.01 (0.002– 0.06) 0.07 (0.03 – 0.16) 0.01 (0.002– 0.06) 0.01 (0.002– 0.06) 0.01 (0.002– 0.06) 0.01 (0.002– 0.06) 0.02 (0.006– 0.08)	0.0033
>60	Lung[ab] Liver[b] Stomach[a] Pyriform fossa[b] Nasopharynx[b] Tongue[b] Colon[b] Bladder[b] Cervix[b] Breast[b] Oesophagus[b]	40(25.64) 8(5.12) 64(41.02) 1(0.64) 2(1.28) 3(1.92) 5(3.20) 7(4.48) 1(0.64) 4(2.56) 19(12.17)	23 5 49 1 1 1 3 2 0 1 9	17 3 15 0 1 2 2 5 1 3 10	0.08 (0.01 – 0.49) 0.18 (0.04 – 0.73) 0.18 (0.04 – 0.73) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49) 0.08 (0.01 – 0.49)	0.2237
2014						
20-30	Oesophagus[ab] Cervix[a] Breast[ab] Colon[b]	2(22.22) 4(44.44) 2(22.22) 1(11.11)	2 0 0 0	0 4 2 1	0.22 (0.05 – 0.91) 1.20 (0.38 – 3.70) 0.10 (0.01 – 0.60) 0.10 (0.01 – 0.60)	0.0265
31-40	Liver[b]	1(2.85)	1	0	0.20 (0.08 – 0.48)	0.0037

	Breast[ab]	7(20.00)	1	6	0.75 (0.38 – 1.44)	
	Colon[ab]	3(8.57)	2	1	0.25 (0.11 – 0.55)	
	Oesophagus[b]	1(2.85)	1	0	0.06 (0.01 – 0.22)	
	Cervix[a]	12(34.28)	0	12	0.02 (0.005– 0.17)	
	Stomach[ab]	9(25.71)	3	6	0.02 (0.005– 0.17)	
	Pyriform fossa[ab]	2(5.71)	1	1	0.06 (0.01 – 0.22)	
	Nasopharynx[b]	1(2.85)	1	0	0.02 (0.005– 0.17)	
41-50	Liver[bc]	7(8.33)	6	1	0.13 (0.06 – 0.25)	<0.0001
	Cervix[a]	22(26.19)	0	22	0.48 (0.30 – 0.75)	
	Oesophagus[ab]	16(19.04)	15	1	0.16 (0.08 – 0.29)	
	Mouth[c]	1(1.19)	0	1	0.07 (0.03 – 0.16)	
	Epiglottis[c]	1(1.19)	0	1	0.22 (0.13 – 0.39)	
	Breast[abc]	10(11.90)	0	10	0.03 (0.01 – 0.10)	
	Stomach[ab]	16(19.04)	13	3	0.03 (0.01 – 0.10)	
	Colon[bc]	3(3.57)	2	1	0.04 (0.01 – 0.12)	
	Larynx[c]	1(1.19)	1	0	0.01 (0.002– 0.06)	
	Pyriform fossa[bc]	3(3.57)	3	0	0.01 (0.002– 0.06)	
	Uterus[bc]	2(2.38)	0	2	0.01 (0.002– 0.06)	
	Penis[c]	1(1.19)	1	0	0.01 (0.002– 0.06)	
51-60	Lung[c]	3(2.80)	2	1	0.06 (0.02 – 0.14)	<0.0001
	Liver[c]	8(7.47)	5	3	0.16 (0.08 – 0.29)	
	Ovary[c]	4(3.73)	1	3	0.36 (0.22 – 0.58)	
	Cervix[bc]	10(9.34)	0	10	0.62 (0.40 – 0.95)	
	Oesophagus[ab]	30(28.03)	26	4	0.03 (0.01 – 0.10)	
	Stomach[a]	31(28.97)	24	7	0.01 (0.002– 0.06)	
	Epiglottis[c]	1(0.93)	1	0	0.01 (0.002– 0.06)	
	Nasopharynx[c]	1(0.93)	0	1	0.02 (0.006– 0.08)	
	Pyriform fossa[c]	5(4.67)	5	0	0.03 (0.01 – 0.10)	
	Uterus[c]	2(1.86)	0	2	0.01 (0.002– 0.06)	
	Gall bladder[c]	2(1.86)	2	0	0.01 (0.002– 0.06)	
	Breast[c]	5(4.67)	0	5	0.01 (0.002– 0.06)	
>60	Lung[ab]	33(25.00)	12	21	0.03 (0.01 – 0.10)	0.0008
	Liver[bc]	10(7.57)	7	3	0.10 (0.05 – 0.19)	
	Stomach[a]	45(34.09)	29	16	0.85 (0.58 – 1.25)	
	Vocal Cord[c]	1(0.75)	1	0	0.01 (0.005– 0.07)	
	Nasopharynx[c]	2(1.51)	2	0	0.15 (0.08 – 0.26)	
	Pyriform fossa[c]	1(0.75)	1	0	0.10 (0.05 – 0.19)	
	Colon[c]	6(4.54)	6	0	0.07 (0.03 – 0.14)	
	Oesophagus[bc]	13(9.84)	12	1	0.03 (0.01 – 0.10)	
	Cervix[c]	8(6.06)	0	8	0.009(0.001-0.05)	
	Breast[c]	4(3.03)	0	4	0.009(0.001-0.05)	
	Epiglottis[c]	2(1.51)	2	0	0.01 (0.005– 0.07)	
	Anus[c]	2(1.51)	0	2	0.009(0.001-0.05)	
	Vaginal Wall[c]	2(1.51)	1	1	0.009(0.001-0.05)	
2015						
20-30	Cervix[a]	5(45.45)	0	5	0.14 (0.03 – 0.56)	0.1132
	Thyroid[b]	2(18.18)	0	2	1.00(0.38 – 2.57)	
	Stomach[b]	1(9.09)	1	0	0.14 (0.03 – 0.56)	
	Breast[b]	2(18.18)	0	2	0.14 (0.03 – 0.56)	
	Ovary[b]	1(9.09)	0	1	0.14 (0.03 – 0.56)	
31-40	Liver[ab]	4(16.00)	4	0	0.34 (0.17 – 0.67)	<0.0001
	Breast[ab]	4(16.00)	0	4	1.15 (0.63 – 2.07)	
	Thyroid[b]	2(8.00)	1	1	0.07 (0.02 – 0.22)	

	Pyriform fossa[b]	1(4.00)	1	0	0.13 (0.05 – 0.32)	
	Bladder[b]	1(4.00)	1	0	0.04 (0.01 – 0.18)	
	Uterus[b]	1(4.00)	0	1	0.02 (0.004– 0.13)	
	Cervix[a]	9(36.00)	0	9	0.02 (0.004– 0.13)	
	Stomach[b]	2(8.00)	1	1	0.02 (0.004– 0.13)	
	Caecum[b]	1(4.00)	1	1	0.04 (0.01 – 0.18)	
41-50	Liver[ab]	7(12.28)	4	3	0.21 (0.13 – 0.34)	<0.0001
	Lung[b]	3(5.26)	2	1	0.38 (0.25 – 0.57)	
	Oesophagus[a]	14(24.56)	14	0	0.33 (0.22 – 0.50)	
	Stomach[ab]	8(14.03)	5	3	0.12 (0.06 – 0.21)	
	Breast[ab]	14(24.56)	0	14	0.02 (0.008– 0.07)	
	Cervix[b]	4(7.01)	0	4	0.008 (0.001-0.04)	
	Endometrium[b]	3(5.26)	0	3	0.04 (0.01 – 0.10)	
	Urinary Bladder[b]	1(1.75)	1	0	0.008 (0.001-0.04)	
	Mouth[b]	1(1.75)	1	0	0.04 (0.01 – 0.10)	
	Epiglottis[b]	1(1.75)	1	0	0.008 (0.001-0.04)	
	Uterus[b]	1(1.75)	0	1	0.01 (0.004 – 0.06	
51-60	Liver[ab]	11(13.92)	4	7	0.06 (0.03 – 0.14)	<0.0001
	Lung[abc]	9(11.39)	4	5	0.18 (0.10 – 0.30)	
	Stomach[a]	15(18.98)	8	7	0.30 (0.20 – 0.47)	
	Breast[bc]	5(6.32)	0	5	0.11 (0.06 – 0.20)	
	Thyroid[bc]	2(2.53)	1	1	0.009(0.001-0.05)	
	Oesophagus[a]	16(20.25)	16	0	0.02 (0.009– 0.08)	
	Pyriform fossa[bc]	4(5.06)	4	0	0.22 (0.13 – 0.35)	
	Gall Bladder[c]	1(1.26)	0	1	0.05 (0.02 – 0.12)	
	Endometrium[bc]	3(3.79)	0	3	0.07 (0.03 – 0.15)	
	Bladder[bc]	2(2.53)	2	0	0.01 (0.005– 0.06)	
	Tongue[c]	1(1.26)	1	0	0.05 (0.02 – 0.12)	
	Cervix[bc]	4(5.06)	0	4	0.009(0.001-0.05)	
	Pharynx[c]	1(1.26)	1	0	0.009(0.001-0.05)	
	Epiglottis[c]	1(1.26)	0	1	0.009(0.001-0.05)	
>60	Liver[bc]	11(11.70)	8	3	0.03 (0.01 – 0.10)	0.2267
	Lung[a]	28(29.78)	15	13	0.64 (0.43 – 0.95)	
	Epiglottis[c]	1(1.06)	1	0	0.10 (0.05 – 0.19)	
	Stomach[ab]	25(26.59)	15	10	0.17 (0.10 – 0.29)	
	Oesophagus[bc]	11(11.70)	8	3	0.15 (0.08 – 0.26)	
	Caecum[c]	1(1.06)	0	1	0.01 (0.005– 0.07)	
	Breast[c]	2(2.12)	0	2	0.02 (0.009– 0.08)	
	Larynx[c]	1(1.06)	1	0	0.01 (0.005– 0.07)	
	Pharynx[c]	1(1.06)	1	0	0.009(0.001–0.05)	
	Bladder[c]	1(1.06)	1	0	0.01 (0.005– 0.07)	
	Colon[c]	4(4.25)	3	1	0.009(0.001-0.05)	
	Pyriform fossa[c]	2(2.12)	2	0	0.009(0.001-0.05)	
	Cervix[c]	3(3.19)	0	3	0.009(0.001-0.05)	
	Ureter[c]	1(1.06)	1	0	0.01 (0.005– 0.07)	
	Gall Bladder[c]	2(2.12)	1	1	0.03 (0.01 – 0.10)	

OR – Odds Ratio CI – Confidence Interval

a, b, c, ab, bc – level of significance using Duncan's ANOVA Test

RESULTS

A total of 1,477 patients with malignancy reported to Genesis Laboratory (Aizawl) from 2011 – 2015. The highest incidence of cancer was observed in the patients with age group 50 - 60 (423), followed by age groups above 60 (399). This shows an increase in the disease with increase in age. There was a statistical significance observed between the

age group and different types of cancers. In 2011, 2012 and 2014, the age group 20-30, 30-40, 40 – 50 and 50 to 60 are shown to be significant, whereas in 2013 and 2014 the age group 30-40, 40-50 and 50-60 shows significance (Table 1). Remarkably, only in 2014 the age group more than 60 years of age was significant with different cancer types. In age group 20-30, breast and cervix cancer were more prevalent, throughout all the years. In middle age group (30-40 and 40-50) cervix, stomach and oesophagus cancer were more prevalent, throughout all the years. In older age group (41 to 60 and >60), stomach cancer was the leading cancer followed by oesophagus cancer in throughout all the years. Cancer is dominant in Females (778) than in males (710). The common type of cancer in female includes cervical cancer, followed by breast cancer which is shown to be increasing with age of the patients and also increasing each year within this 5 year period from 2011 – 2015 (Figure 1). In males, stomach cancer is the most prevalent type of cancer followed by oesophagus cancer. Both the type of cancers are found commonly in the patients with age 40 yrs and above.

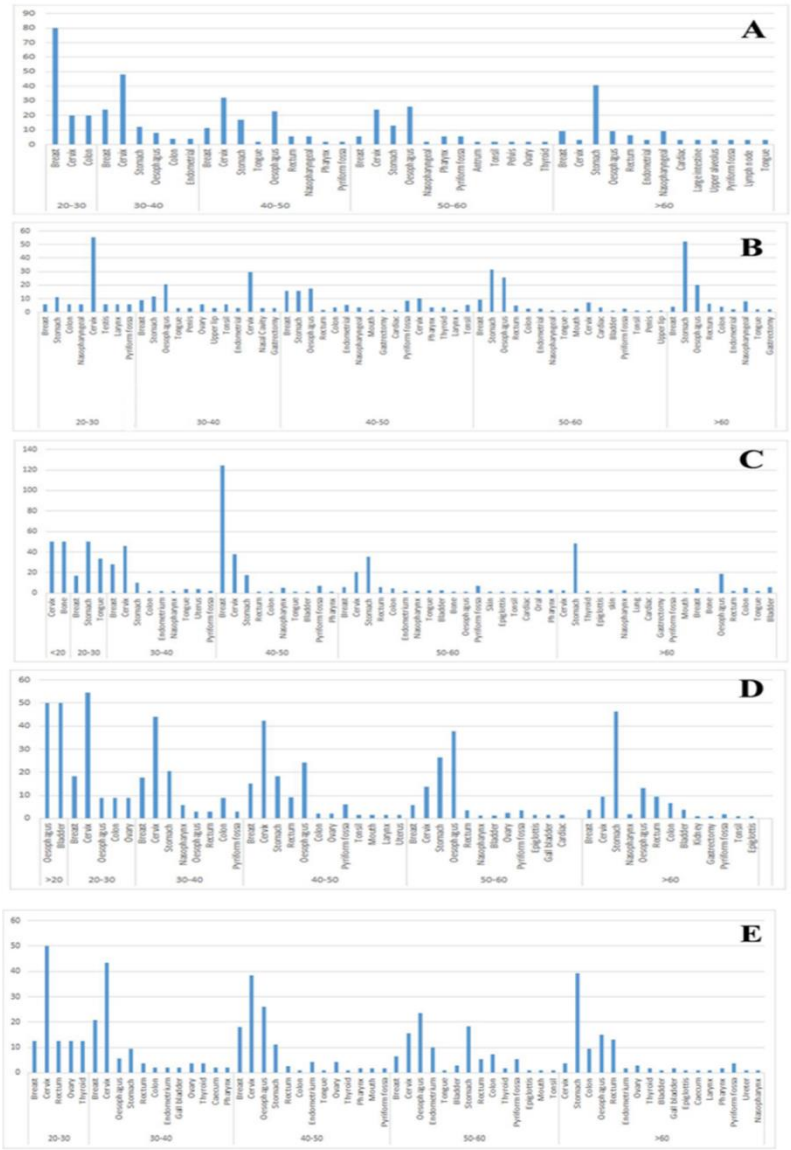
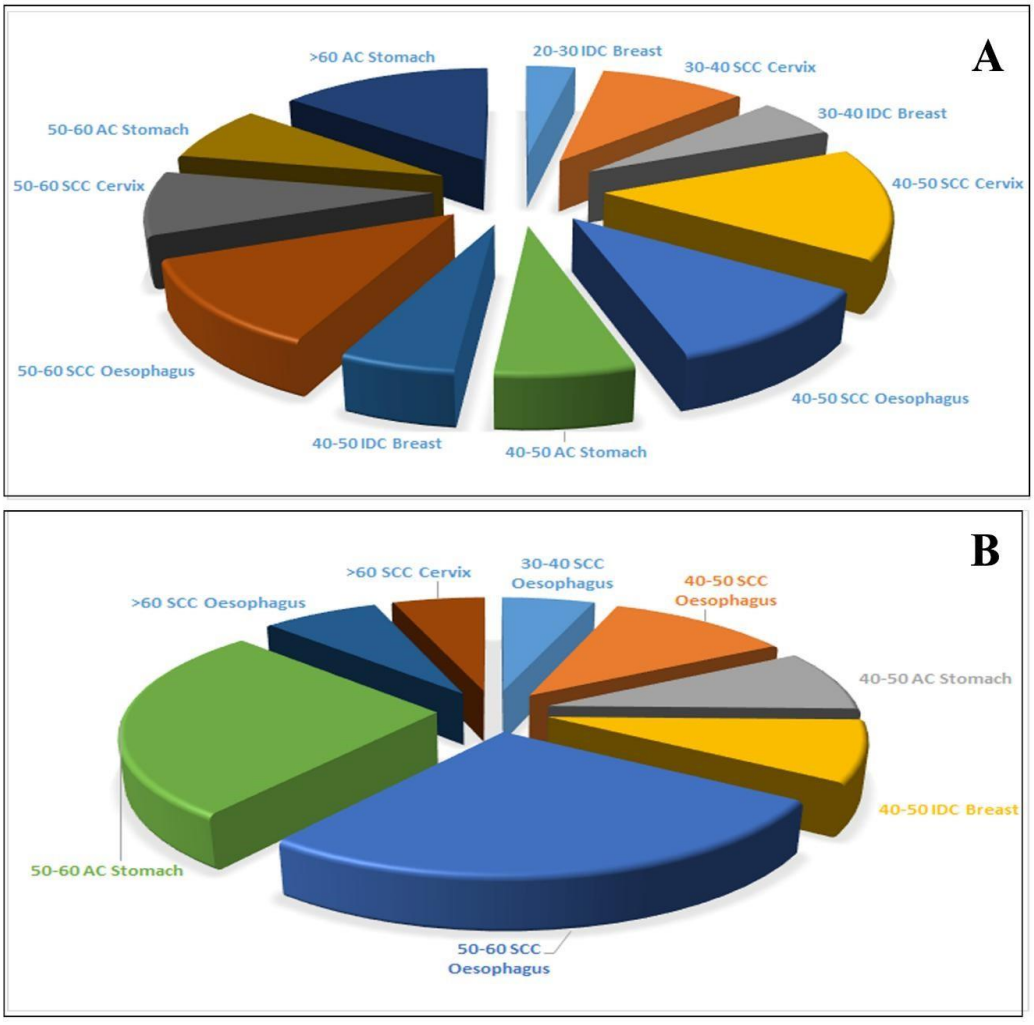
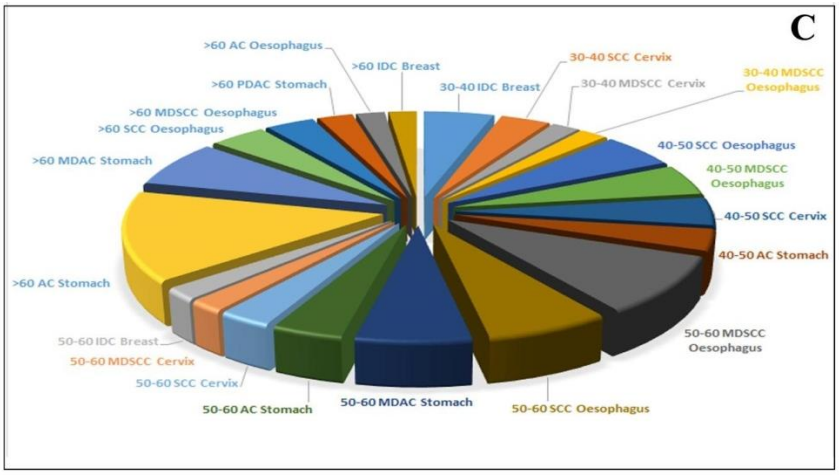
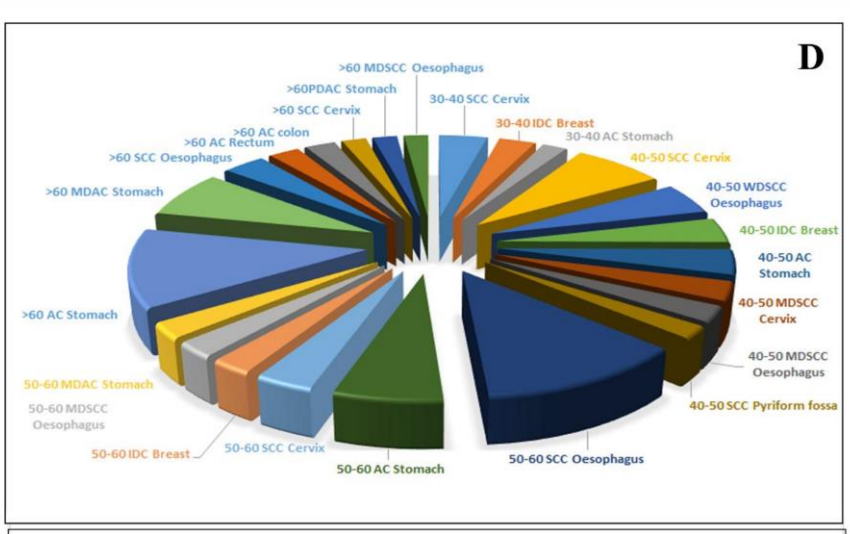


Figure 1. Various Cancer Frequency based on Age group for the years 2011 (A), 2012 (B), 2013 (C), 2014 (D), 2015 (E).

Different types of cells were involved with the different types of cancers. In 2011, Squamous cell carcinoma (SCC) was the most common pathological variety of cancer in males, followed by Adenocarcinoma (AC). In 2013, Moderately Differentiated Squamous Cell Carcinoma (MDSCC) had the highest percentage of occurrence. In 2014 and 2015 AC and SCC are both commonly seen. In the year 2011 the adenocarcinoma stomach and squamous cell carcinoma oesophagus showed prevalence in older age group, but in middle age group the squamous cell carcinoma cervix was the leading cancer group (Figure 2). Adenocarcinoma stomach and squamous cell carcinoma oesophagus were the ladder of all the cancer types for the year of 2012. In 2013 the older age group had significant growth than other age group, because the adenocarcinoma stomach and moderately differentiated squamous cell carcinoma in oesophagus were the leading cancer. The same prevalence was also observed for the year of 2014 and 2015. But for 2015 other types of cancers prevalence was also observed.







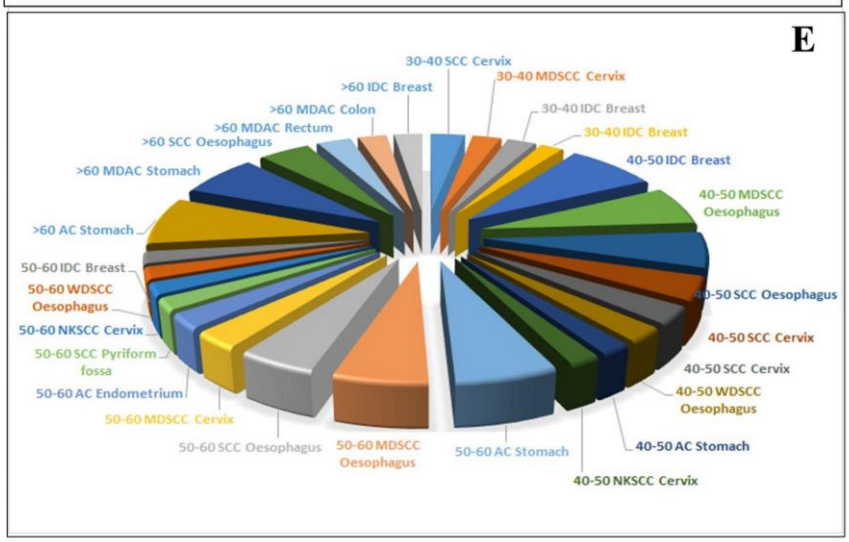


Figure 2. Frequency of cancer prevalence and status in Mizo population for the years 2011 (A), 2012 (B), 2013 (C), 2014 (D), 2015 (E).

AC – Adenocarcinoma; IDC – Invasive Duct Carcinoma; SCC – Squamous Cell Carcinoma; MDAC – Moderately Differentiated Adenocarcinoma; MDSCC – Moderately Differentiated Squamous Cell Carcinoma; PDAC – Poorly Differentiated Adenocarcinoma; WDSCC – Well Differentiated Squamous Cell Carcinoma; NKSCC – Non – Keratinising Squamous cell carcinoma.

DISCUSSION

Our studies reveal that Cancer is diagnosed more in women than in men. It also shows an increase in occurrence of cancer with increase in age. Women are showing high rate of cancer because of the alarming increase in cervical and breast cancer for the last five years. Cervical cancer is caused by human papillomavirus infection. Human papillomavirus promotes uncontrolled cell division and accumulation of genetic damage. Other cancers attributed to human papillomavirus infection include those of vagina (70%), penis (50%), vulva (43%), and oropharynx (26%) (Table 1)[Hariri et al.,2011]. Breast cancer occurrence is highest after 35 years of age. The possible risk factor for occurrence of breast cancer was contributed by stress in the form of higher education and occupation, late menopause, history of induced abortion, first-degree family history of the disease and body mass index [Thapa et al.,2016]. The most common type in males is stomach cancer, followed by oesophageal. Tobacco smoking and use of smokeless tobacco, chewing tobacco and tuibur coupled with unhealthy food items such as smoked meat and vegetables and fermented soyabean and pork etc. may attribute to the high incidence of cancer in Mizoram [Ghatak et al.,2016].

The food habits of Mizoram are known to contribute to various different types of cancers. Smoke-drying and preservation leads to formation of N-nitroso compounds. Nitrite reacts with amines and amides found in meats and other proteins to form N-nitroso compounds, which are animal carcinogens and possible human carcinogens. Furthermore, although salt is not a carcinogen, it is thought to increase the risk of gastric cancer through direct damage to the gastric mucosa, which results in gastritis, increased DNA synthesis, and cell proliferation. This indirectly contributes to the development of chronic atrophic gastritis, leading to the development of stomach cancer. Because of the presence of both salt and nitrite in processed fish and meats, its role in the development of stomach cancer cannot be ignored, as was found by Phukan et al. [Phukan et al.,2006]. Frequent consumption of sa-um was found to be associated with the risk of developing stomach cancer. This is a food material uniquely consumed in Mizoram. Dietary intakes of total or saturated fat have been shown to be associated with stomach cancer. Boiled pork fat, in addition to being a rich source of saturated fat, may form carcinogenic compounds during long storage, as in other stored meats. Use of soda was shown to be a risk factor. Indigenous people of the northeastern region of India use soda (alkali) or other alkaline preparations frequently as food additives.

The consumption of tobacco is the leading cause of cancers in India. The regular use of tobacco via smoking, chewing, snuffing etc. in Mizoram, which is responsible for 65 to 85% cancer incidences in men and women, respectively. The various cancers produced by the use of tobacco are of oral cavity, pharynx, esophagus, larynx, lungs and urinary bladder. Smoking is the most notorious factor for the causation of lung cancer [Hammond et al.,1966]. Approximately, 87 and 85% males and females have been found to have lung cancer due to tobacco smoking in the form of local Zozial (a thin South Asian cigarette type structure filled with tobacco flake and wrapped in a white paper, tied with a string at one end) [Behera et al.,2004] and cigarette in India [Jayant et al.,1991]. The severe carcinogenic nature of Asian local made cigarette has been proved by the studies of Jussawalla and Jain [Jussawalla et al.,1979] and [Pakhale et al.,1990]They observed that the unrefined form of tobacco used in bidis (WHO, 1999) and the

frequency with which a bidi needs to be puffed per minute may be responsible for its relatively higher carcinogenic effects as compared to cigarettes [Bano et al.,2009]. Bidi smoking at two puffs per minute produces about equal amounts of carcinogens (steam volatile phenols, hydrogen cyanide and benzopyrene) as produced by one puff per minute of unfiltered cigarette [Pakhale et al.,1990] Hookah (a special cigar used in Mizoram using raw tobacco) smoking causes lung cancer; as reported by Nafae et al. [Nafae et al.,1973] In Mizoram, North-eastern India high incidences of stomach cancer are attributed to the consumption of smoked meat and chewing of tobacco. High incidences of stomach cancer in Mizoram are the result of the excessive use of tuibur (water filterate of tobacco).

Similarly, the consumption of Areca nut, Pan Masala, Opium and Bhang (leaves and flower powder of female cannabis plant) has been recognized as the major cause of mouth cancer in Mizoram. The daily consumption of the number betel leaves by an individual is about 15-25 in various districts of Mizoram, which continuously acts as an irritant to the buccal mucosa [Mehrotra et al.,2003]. One of the most important factors responsible for the oropharyngeal malignancy in Mizoram is the chewing of raw betel nut [Wahi et al.,1965]. Among various risk factors for the occurrence of oesophageal cancer in Mizoram, betel quid chewing carries a relative risk of 1.5 to 3.5%. The salted cooked vegetables made by adding sodium bicarbonate has shown to possess a high methylation activity and may lead to the endogenous formation of nitrosamine [Malkan et al.,1997] and that can lead to stomach cancer.

Alcohol consumption has been considered as one of the major causes of colorectal cancer as per a recent monograph of WHO [Baan et al.,2007]. Annually, about 9.4% new colorectal cancer cases are attributed to the consumption of alcohol, globally [Parkin et al.,2006]. An increased risk of 10% was observed with consumption of more than two drinks per day, which suggests a causative role of alcohol consumption in colorectal cancer [Toriola et al.,2008]. Recently, a study revealed that an increased risk of colorectal cancer was limited to consumption of more than 30.0 g of alcohol per day [Longnecker et al.,1990]. Relationship between alcohol consumption and high risk of oesophageal cancer was first known in 1910 [Tuyns et al.,1979]. However, chronic alcohol consumption has been found to be a risk factor for the cancers of the upper respiratory and digestive tracts, including oral cavity, hypopharynx, larynx and oesophagus as well as liver, pancreas, mouth and breast cancers [Tuyns,1979;Maier,1994;Seitz et al.,2004]. A 10.0 g/day intake of alcohol by a woman increases its relative risk of breast cancer by 7.1% [Doll et al.,1981]. The mechanism of carcinogenesis due to alcohol consumption is not exactly known, however, it is thought that ethanol being a co-carcinogen might play a crucial role in the carcinogenesis [Poschl et al.,2004]. The metabolic products of ethanol are acetaldehyde and free radicals. The free radicals are responsible for alcohol assisted carcinogenesis through their binding to DNA and proteins, which destroy foliate leading to secondary hyper proliferation [Anand et al.,2008].

Non–tobacco risk factor includes infections, dietary factors, alcohol use, physical activities and body composition. Other risk factors include exposure to asbestos, air pollution (indoor and outdoor), occupational exposures and exposure to radiation. Consumption of alcohol is shown to be associated with cancers of the mouth, pharynx, larynx, oesophagus, colo-rectum (men) and breast (pre- and post- menopausal). Aflatoxins causes liver cancer. Arsenic in drinking water and beta-carotene supplements are known to contribute to lung cancer. Colo-rectum is known to be caused by excess consumption of red and processed meat in our diet. Associations between infections and cancer: HPV(Human Papilloma Virus) infection and cervical cancer, Epstein Barr Virus(EBV) and Burkitt Lymphoma as well as Non-Hodgkin and Hodgkin Lymphoma, Hepatitis-C virus and hepatocellular

carcinoma, Kaposi Sarcoma Herpes Virus (KSHV) and Kaposi Sarcoma [Harford et al.,2012].

Acknowledgements

The authors thank Bioinformatics Infrastructure (BT/BI/12/060/2012 (NERBIF-MUA), and DBT-DeLCON Facility at Mizoram University sponsored by the Department of Biotechnology (DBT), New Delhi, Govt. of India. The authors thank Mary K Zothanpari for her help during data collection.

Conflict of interest

The authors declare that no competing or conflict of interests exists. The funders had no role in study design, writing of the manuscript, or decision to publish.

Authors' contributions

JZ and NSK conceptualized the study; MVT and FL collected the primary data; MVT, SG and FL did the analysis and interpretation of the data; MVT, SG, NSK and JZ wrote the manuscript.

REFERENCES

Anand, P., Ajaikumar, B.K., Sundaram, C., Kuzhuvelil ,B.H., Sheeja ,T.T., Oiki, S.L., Sung ,B., and Bharat ,B.A.(2008). Cancer is a Preventable Disease that Requires Major Lifestyle Changes. Pharm Res ; 25: 2097-2116.

Baan ,R., Straif, K .,and Grosse, Y.(2007). Carcinogenicity of alcoholic beverages. Lancet Oncol ; 8: 292-3.

Bano, R., Mahagaonkar ,A.M., Kulkarni ,N.B., Ahmad, N., Nighute ,S.(2009). A study of pulmonary function tests among smokers and non-smokers in a rural area. Pravara Med Rev ; 4: 11-16.

Behera, D., Balamugesh, T .(2004). Lung Cancer in India. Indian J Chest Dis Allied Sci 6, 269-281.

Bener, A., Ayub, H., Kaki,R., Wanis.(2008). Ibrahim Patterns of Cancer Incidence Among the Population of Qatar: A Worldwide Comparative Study. Asian Pacific J Cancer Prev ; 9:19-24.

Doll ,R., Peto ,R. (1981). The causes of cancer: quantitative estimates of avoidable risks of Cancer in the United States today. J Natl Cancer Inst 66, 1191-308.

Ghatak, S., Yadav, R.P., Lalrohlui, F.(2016). Xenobiotic Pathway Gene Polymorphisms Associated with Gastric Cancer in High Risk Mizo-Mongoloid Population, Northeast India. Helicobacter; 21: 523–535.

Hammond ,E.C.(1966). Smoking in relation to the death rates of one million men and women. In: Haenszel W, editor. Epidemiological Approaches to the Study of Cancer and Other Chronic Diseases. National Cancer Institute Monograph No 19. Bethesda (MD): U.S. Department of Health, Education, and Welfare, Public Health Service, National Cancer Institute; pp: 127–204.

Harford ,J.B.(2012). Viral infections and human cancers: the legacy of Denis Burkitt. British Journal of Haematology ; 156: 709–718.

Hariri, S., Dunne, E., Saraiya, M., Unger, E., and Markowitz, L.(2011). Chapter 5: Human Papillomavirus. VPD Surveillance Manual ; 5: 1-5.

Jayant, K., Notani, P.N., Gulati ,S.S., Gadre ,V.V.(1991). Tobacco use in school children in Bombay, India. A study of knowledge, attitude and practice. Indian Journal of Cancer ; 28:139–147.

Jussawalla, DJ., Jain, DK.(1979). Lung cancer in greater Bombay correlation with religion and smoking habits. Br J Cancer ; 40: 437-48.

Longnecker, M.P., Orza, M.J., Adams, M.E., Vioque, J and Chalmers, T.C.(1990). A meta-analysis of alcoholic beverage consumption in relation to risk of colorectal cancer. Cancer Cause Control; 1: 59-68.

Maier ,H., Sennewald ,E., Heller, G.F., and Weidauer, H.(1994). Chronic alcohol consumption The key risk factor for pharyngeal cancer. Otolaryngol Head Neck Surg ;110: 168-73

Malkan G and Mohandas KM.(1997). Epidemiology of digestive cancers in India. I. General principles and oesophageal cancer. Indian J Gasteroenterol ; 16:98-102.

Mehrotra ,R., Singh ,M., Kumar, D., Anpandey, Gupta ,R.K and Sinha, U.S.(2003). Age specific Incidence rate and pathological spectrum of Oral cancer in Allahabad. Ind J Med Sci; 57, 400-404.

Murthy ,V.H., Krumholz, H.M., and Gross, C.P.(2004). Participation in Cancer Clinical Trials Race-, Sex-, and Age-Based Disparities. JAMA ; 291: 2720-2726.

Nafae, A., Misra ,S.P., Dhar, S.N., Shah S.N.(1973).A Bronchogenic carcinoma in Kashmir valley. Indian J Chest Dis 1973; 15: 285 - 95.

National Cancer Registry Programme (NCRP).(2010). Three year report of the population based cancer registries 2006-2008. (First report of 20 PBCRs in India), Bangalore, Indian Council Med Res ; 11:71–8.

Pakhale ,S.S., Jayant ,K., Bhide ,S.V.(1990) Chemical analysis of smoke of Indian cigarettes, bidis and other indigenous forms of smoking, levels of phenol, hydrogen cyanide and benzopyrene. Indian J Chest Dis Allied Sci ; 32: 75-81.

Parkin, D., Bray ,F., Ferlay, J.(2006). Global cancer statistics, 2002. CA Cancer J Clin; 55:74–108.

Phukan ,R.K ., Narain ,K., Zomawia, E., Hazarika N.C., and Mahanta ,J.(2006). Dietary habits and stomach cancer in Mizoram. India. J Gastroenterol ; 41: 418–424.

Poschl ,G., Seitz ,H.K.(2004). Alcohol and cancer. Alcohol ; 39: 155-65.

Rao, C.V., Newmark ,H.L .,and Reddy ,B.S.(1998). Chemopreventive effect of squalene on colon cancer. Carcinogenesis 1998; 19: 287–290.

Seitz ,H.K., Stickel, F., and Homann ,N.(2004). Pathogenetic mechanisms of upper aerodigestive tract cancer in alcoholics. Int. J Cancer ; 108, 483-7.

Toriola ,A.T., Kurl, S., Laukanen, J.A., Mazengo, C., Kauhanen J.(2008). Alcohol consumption and risk of colorectal cancer: The Findrink study. Eur J Epidemiol ; 23: 395-401.

Thapa, S., Lalrohlui, F., Ghatak, S., Zohmingthanga, J., Lallawmzuali, D., Pautu, J.L .,and Kumar ,N.S.(2016). Mitochondrial complex I and V gene polymorphisms associated with breast cancer in mizo-mongloid population. Breast Cancer; 38: 1-5.

Tuyns ,A.J.(1979). Epidemiology of alcohol and cancer. Cancer Res ; 39: 2840–3.

Wahi ,P.N., Kehar, U., and Lahiri, B.(1965). Factors influencing oral and oropharyngeal cancers in India. Brit J Cancer ; 19: 642-60.

PARTICULARS OF CANDIDATE

NAME OF CANDIDATE : Freda Lalrohlui

DEGREE : Ph.D

DEPARTMENT :Biotechnology

TITLE OF THESIS : Genetic variation studies in candidate genes related to Type 2 Diabetes in Mizo population, Mizoram

DATE OF ADMISSION : 11.08.2015

APPROVAL OF RESEARCH PROPOSAL:

1.BOARD OF STUDIES : 7.04.2016

2.SCHOOL BOARD : 22.04.2016

REGISTRATION NO &DATE : MZU/Ph.D/ 841 of 22.04.2016

EXTENSION(IF ANY) : NA

BIO-DATA

Name	: FREDA LALROHLUI
Father's Name	: JOHN LALREMRUATA
Gender	: FEMALE
Address	: C-21/1, RAMTHAR VENG, AIZAWL, Pin-796001
Subject of Specialization	:EPIDEMIOLOGY AND MOLECULAR GENETICS IN DIABETES
Education Qualification	: M.SC(BIOTECHNOLOGY)GOLD MEDALIST
Marital Status	: SINGLE
Number of publications	: 9 publications in peer reviewed journals