Introduction & Review of literature

Tuberculosis commonly (TB) is caused by a single infectious agent, *Mycobacterium tuberculosis (Mtb)* (Tong *et al.,* 2017). TB spreads from person who is diseased to other healthy person through the air droplets generated by cough, sneeze or spit, and become infected. An individual is believed to have a maximum of 15% chance of becoming sick with TB, this is because most of the people has latent TB and thus being unable to spread the disease. The probability of infection with the disease is higher in immunocompromised people. These individuals can be, people with HIV and diabetes, people who are undernourished, people with the habit of smoking or use of tobacco and alcohol consumption. The early symptoms of TB disease include coughing, timely fever, night sweats, and the person also loses weight because they have no appetite for food for more than two weeks. Since the symptoms may be mild for months, it may lead to delayed detection and result in disease transmission. It has been estimated that, people suffering from TB, on close contact can infect approximately 15 other healthy person. Therefore, early detection and proper treatment is needed to reduce TB burden (Global Tuberculosis Report, 2019).

The host genetic components are important actors in regulating the progression of the disease, and 10% of the infected people develop the active disease. The cell mediated immune responses determines and controls the progression of the infection, from latent to the active pulmonary or extra-pulmonary tuberculosis (Tong *et al.,* 2017). Differences in occurrence, outcomes and rates of TB among different races, ethnicities and even within the family shows the contribution of genetics towards tuberculosis susceptibility, which shows the

complexity of *Mtb* infection where the host, environmental and the bacteria play a critical role. Several genomic studies demonstrate that susceptibility to tuberculosis is strongly influenced by the host genetics (Moller *et al.,* 2018). Understanding these genetic variations may contribute and pave way for understanding the mode of action of the pathogenesis of tuberculosis, contributing to the development of strategies which can help in treatment of patients and also prevention of the disease.

Global and National TB disease burden

Tuberculosis is commonly found in every country. New incidences of TB, is largely reported in regions of South-East Asia and Western Pacific regions contributing upto 62%, which is followed by 25% in regions of Africa (WHO, 2018). World Health Organization (2018) reports the occurrence of new TB cases in the thirty high TB burden countries to be 87%. China, Indonesia, India, Bangladesh, Pakistan, South Africa, Nigeria and the Philippines are the top eight countries with high incidences of TB, accounting for about two thirds of the new TB cases. Globally, an estimated 10 million individual are developing the disease and out of these the diseased men, women and children are 5.8 million, 3.2 million and 1 million, respectively. The number of deaths caused by this infectious disease alone is estimated to be 1.3 million although the incidence rate is falling at the rate of 2% annually, which needs to reach about 5% to accomplish the 2020 milestones of the end TB strategy.

In India, as per the TB report 2018, the incidence of TB was 2.7 lakhs approximately, which is about a quarter of the global TB cases. The current statistics of different forms of TB cases and mortality rate are listed in Table1.

Table 1: Estimates of TB Burden in India, 2018

Indicators	No. of people infected
Incidence of TB (including HIV)	27,90,000
Mortality due to TB (Excluding HIV)	4,23,000
Incidence of MDRTB/RR	1,47,000
Incidence of HIV-TB	87,000
Mortality due to HIV-TB co-morbidity	12,000

Mizoram is a small hilly tribal state situated in the north eastern region of India, bounded by Cachar district of Assam in the north, Manipur to the north-east, Bangladesh to the south-west and Myanmar to the east and south. And the people of Mizoram are called as “Mizo”. Mizoram became a state in February 1987. Mizoram is a house for 1,097,206 people, with 555,339 males and 541,867 females as per 2011 census. Mizoram contributes only about 0.09% of the population in India and 50% of urban population is living in the capital city, Aizawl. The Mizo’s are categorised as Tribal which is again divided into sub-tribes constituting of about 96.3% of population in Mizoram. The main occupation of the people is agriculture on small land holdings and Jhum cultivation, livestock and dairy farming and agro forest produce are also a minor trade of the region.

Various tribes migrated from different places, arrived in waves and made these hills their homeland. All the tribes of Mizoram actively participate in Hnatlang, or community service is an important activity for community welfare. Above all, most Mizo tribes believed in the ethics of Tlawmngaihna moral code which made them courteous and helpful to others.The prevalence rate of TB disease in Mizoram is much higher than the national average, which is 1063 per 100,000 population. At

least 2,499 people were diagnosed with TB in 2018. Of the suspected 1,685 MDR-TB (multi-drug-resistant tuberculosis) patients, 21 were confirmed to be diagnosed with the disease, while 123 people suffered from both TB and HIV positive. In 2017, as many as 1,931 people had been diagnosed with TB and 2,149 people in 2016. This shows that a massive effort is needed to fight TB in Mizoram (TB India Report 2018).

Risk factors for TB

Persons with weak immune system have a much higher risk of developing the disease than those with a stronger immune system that is capable of fighting back the pathogen. Also person recently infected with the bacteria and with medical conditions having other malignancy is a risk factor for TB.

Studies report the risk of TB infection to be ten times higher in HIV infected person, which shows the importance of TB treatment in HIV patients (Menzies *et al.,* 2008). According to the WHO's guidelines, the treatment needs to be stressed in both developed and developing countries. An individual with diabetes is also known to have a higher risk of TB infection, and the relative risk is reported to range from 1.16-7.83 (Harries *et al.,* 2011; Dobler *et al.,* 2012). Malnutrition leads to being under-weight which can cause about two fold increase in development of TB compared to the normal or overweight individual (Palmer *et al.,* 1957). People who do not get enough healthy food or food to eat are at a risk of TB. Lifestyle factors and food habits may also be a risk factor contributing to the development of the disease (Abba *et al.,* 2008). The person with the habit of smoking are at a high risk of TB as smoking alters immune responses to the bacterium as a result of which it

contributes to susceptibility to TB infection (Chan *et al.,* 2014). TB infection are higher in tobacco smokers as compared to non-smokers (Slama *et al.,* 2007). It has been estimated that reduction in TB risk will be achieved by completely eliminating tobacco smoking (Smit *et al.,* 2010).The consumption of alcohol impairs the immune system and the functioning of alveolar macrophage in eliminating the pathogens get compromised, which increases the susceptibility to TB (Imtiaz *et al.,* 2017). A systematic review concluded that people who drink alcohol everyday or have an alcohol use disorder, are at a higher risk of active tuberculosis (Lönnroth *et al.,* 2008). Health care workers, who are known to be in close contact with patients and aboriginal population are also at an increased risk of TB infection and disease. A plan of action for early detection of people in need of TB treatment may also be achieved by identifying the risk factors for the development of the disease.

Role of host genetics

Tuberculosis was thought to be hereditary, because it was observed that it frequently occurred in families living together (Moller *et al.,* 2009). Several genetic variants associated with tuberculosis infection are population specific. Therefore, population specific studies are required to understand the underlying genetics of susceptibility to tuberculosis, determining immunological response to infection, and to understand target genes for treatment success and efficacy of vaccine. Clinical and epidemiological observations of TB susceptibility supported the involvement of human genetics in TB susceptibility. Qu'Appelle Indians, who did not have previous exposure to TB, showed a mortality rate of 10%. Eventually, the death rate decreased to 0.2% after years of exposure (Stead *et al.,* 1990). European individuals and Ashkenazi Jews are less susceptible to the disease as they are in

contact with the bacterium for centuries in Europe, resulting in the selection of more resistant population. This difference between populations was not due to socioeconomic factors, but the genetic makeup of the individual (Motulsky *et al.,* 1960). Thus, there might be existence of inherited genetic susceptibility for a host defense against TB development.

Candidate gene study is dependent on prior knowledge, and is capable of detecting genes which are known to have a limited effect. Solute carrier family 11 member 1 gene (SLC11A1), situated at Chr2q35, is the first and amongst the commonly studied gene with respect to TB.It is 14 kb in length with 15 exons (Qu *et al.,* 2011). SLC11A1 genes are expressed in reticuloendothelial cells. It functions as an exchanger for proton and divalent cation, which alters the phagolysosomal environment. SLC11A1 plays a role in the innate immune response to *Mtb*, lessening DNA replication and respiratory chain function of the bacteria (Courville *et al.,* 2000). Several independent studies have been conducted across different populations between the polymorphisms commonly found in SLC11A1 (3'UTR, D543N, INT4, and 5'[GT]n) to show their association with TB susceptibility (Bellamy *et al.,* 1998; Liu *et al.,* 2004; Duan *et al.,* 2003;Gao *et al.,* 2000; Ryu *et al.,* 2000; Kim *et al.,* 2003). To show association between SLC11A1 and TB susceptibility, a number of meta-analysis have been conducted particularly in Asian and African populations (Li *et al.,* 2006; Li *et al.,* 2011; Meilang *et al.,* 2012).

The Histocompatibility leucocyte antigen genes commonly known as HLA, is another candidate gene commonly studies. This gene is located at Chr6p21.3 and plays an essential role in defence against various infectious diseases. The class II group of HLA is consisting of three regions - DR,DQ and DPmolecules whichpresent

antigenic peptidesto CD4+T cells on the surfaces of antigen-presenting cells (APCs); On the other hand, the class I group of HLA consists of regions such as-A,B,Cmolecules which functions to activate cytotoxic Tcells by presenting foreign peptides on the surfaces of the infected cells (North and Jung, 2004). HLA class II polymorphisms association with TB susceptibility or resistancehave been demonstrated by several studies, with conflicting results. HLA-DRB1 which forms a part of the HLA class II molecules expressed on the surface of antigen presenting cells, plays its role in responding to *Mtb* infection (Kinnear *et al.,* 2017). HLA-B*35:19/47 and HLA-C*03 gene's association with TB susceptibility was shown in a group of people with a shared characteristic of north-eastern Argentina (Habegger *et al.,* 2014). Also in a cohort study of 759 South African Coloured Cases and ethnically matched controls, HLA-B*58 and HLA-Cw10 were contributing to thedevelopment of the disease, while HLA-B*47 and HLACw5 were protecting the individual from developing the disease (Salie *et al.,* 2015). Pulmonary TB has been associated with HLA-DRB1*04, and its subtypes DRB1*04:07:01, DRB1*04:11:01 and DRB1*04:92, could be a potential immunogenic marker involved in disease development (Lima *et al.,* 2016). Indian studies on HLA association with TB shows that DRB1*1501 of HLA-DR2, HLA-DQB1*0601 and DPB1*02 were positively linked with susceptibility to pulmonary tuberculosis (Ravikumar *et al.,* 1999; Uma *et al.,* 2001).

Toll-like receptors genes, commonly known by its acronym TLR are a class of pattern recognition molecules, functions to activate innate immunity against microbial infection inducing expression of genes involved in inflammatory process (Akira *et al.,* 2006). There are ten functional TLRs (TLR1-10) expressed by different cells. TLR genes are known to be polymorphic and diverse among populations. A

partial explanation for male bias in TB ratios can be explained by studying TLR genes, as some of these genes are located in X- Chromosomes. A candidate gene study in Indonesian and Russian population detected some sex specific associations for TLR8 polymorphisms (Davila *et al.,* 2008; Bellamy *et al.,* 2000). An observational study between cases and controls in South African population found that the presence of SNP rs3761624A/G in TLR8 was responsible for susceptibility to TBin females and SNPs rs3764879C/G and rs5743618G/T in TLR1 and rs3764880A/G in TLR8 was responsible for making both males and females susceptible to TB infection (Salie *et al.,* 2015). Also another polymorphism in TLR8, SNP rs3764880A/G (Met1Val), and in TLR1, rs4833095G/A (Ser248Asn) were associated with TB protection, affecting TLR structure, capable of killing the pathogen, *Mtb* (Dittrich *et al.,* 2015). Polymorphisms in TLR2 also showed association to tuberculosis meningitis rather than pulmonary TB in Chinese population (Zhao *et al.,* 2015). TLR1, TLR2, TLR4, TLR6, TLR8, TLR9 and the adaptor molecule TIRAP were screened for populations such as Ghana (Meyer *et al.,* 2016), Chinese (Wu *et al.,* 2015), Iranian (Jafari *et al.,* 2016) and Croatian(Kardum *et al.,* 2015).

Vitamin D receptor (VDR) is a nuclear hormone receptor expressed on the surface of monocytes and activated T and B lymphocytes. Metabolism of Vitamin D restricts the intracellular growth of *Mtb* by activating macrophages. The commonly investigated polymorphisms in VDR are ApaI, TaqI, BsmI and FokI which are significantly associated with susceptibility to TB for different populations such as Han Taiwanese(Lee *et al.,* 2016), Asian (Huang *et al.,* 2015), Caucasian and African (Cao *et al.,* 2016).

Interferon-gamma (IFN-ɣ) is an important member of the INF family whose role is to protect the body from infection. T helper 1 (Th1) response, which is required to contain *Mtb* is characterised by the production of IFN-ɣ by Natural Killer (NK) cells (Moller *et al.,* 2010). Many scientist have studied the association of IFN-ɣ genelocated in the first intron (1874T/A) with variation in DNA. The meta-analysis of 11 studies suggested the relation of IFNG+874 polymorphism with TB susceptibility (Pacheco *et al.,* 2008). Genotyping of IFN-ɣ variants in Korean population showed its association with the risk of TB. The variants were rs9376269C/G, rs9376268A/G, rs9376267C/T and rs56251346C/T (Shin *et al.,* 2015). Genes such as IFN-ɣ,IFN-ɣ receptor1, IFN-ɣ receptor2, Interleukine-12A, Interleukine-12B, Interleukine-12RB1, Interleukine-12RB2, Interleukine-23A, Interleukine-23B, Interleukine-27, and Interleukine-27RA whose proteins contributes to the signalling of IFNG gene were analysed, which shows a significant role or contribution of the variants in protection to TB in Ghanaian population (Meyer *et al.,* 2016).

Interleukin-10 (IL-10) associated with macrophage deactivation is a major anti-inflammatory cytokine and a polymorphism has been suggested to play a role in the immunity and progression of inflammation (Shaw *et al.,* 2000). IL-10 gene polymorphism at loci 1082, 819, and 592 shows association with susceptibility to TB (Ates *et al.,* 2008). Interleukine-6 (IL-6), is produced at the site where it is infected by the bacteria at a very early stage. These cytokines are a macrophage derived pro-inflammatory cytokine thought to be involved in changingitsresting state to being inaction for the clinical outcome of the disease. IL-6 association with the pathogenesis of tuberculosis has also been reported (Lyadova *et al.,* 2010). A variation in IL-6 in the position -174G/C and SNP rs763780C/T in IL-17A and IL-17F was studied to be related with the risk of TB disease in Chinese population (Du *et*

al., 2015) as well as in Asian population (Zhao *et al.,* 2016), but was not reported for Croatian (Kardum *et al.,* 2015) and Caucasians populations (Zhao *et al.,* 2016). These report shows that some populations are genetically predisposed to TB while others are not.

TNF (Tumor Necrosis Factor) -α is an important component of the antimycobacterial cytokine cascade. TNF-α has been suggested to take part in the activation of macrophages and formation of granulomas. Polymorphism in -308 position with A/G substitution is the commonly observed SNP in this gene (Jafari *et al.,* 2016). A meta-analysis performed evaluated potential associations of variants, -308G/A and-238G/A with pulmonary TB in Asian and African individual (Yi *et al.,* 2015). MBL (Mannose-binding lectin), plays major part in protection against *Mtb* by acting as an opsonin and also promotes inflammation and releases cytokines. MBL genes and HLA-DR2 association with TB disease is reported in the population of Indians (Selvaraj *et al.,* 1999).NOS (Nitric oxide synthase), through bacteriostatic processesplays its role in response to *Mtb*, expressed by mycobacterium-infected macrophages (Azad *et al.,* 2012). An analysis of nitric oxide synthase 2A (NOS2A) gene found nine single nucleotide polymorphisms (SNPs) that are shown to be significantly related to the outcome of TB ($P < 0.05$), in a case-control study of African-Americans and Caucasians, multiple SNP interactions between NOS2A and the interferon gamma receptor 1 (IFNGR1) gene (P ranging from 0.0004 to 0.0006), and interactions between NOS2A and the Toll-like receptor-4 (TLR4) gene ($0.002 < P < 0.005$), in the African-American individuals (Velez *et al.,* 2009).

Genome-wide linkage analysis (GWLA) has been used to identify susceptibility loci associated with TB. This analysis identified various genomic regions that showed evidence of linkage to TB such as Chr15q11-q13 (LOD = 2.00)

(Shaw *et al.,* 1997), ChrXp26 (LOD = 1.7) (Bellamy *et al.,* 2000) , Chr15q11-13, Chr2q35(LOD = 3.81) (Cervino *et al.,* 2002),Chr7p22-p21, Chr8q12-q13 (LOD = 3.49) (Greenwood *et al.,* 2000), Chr5q (LOD = 2.29) (Stein *et al.,* 2008), Chr17q13.3-13.1 (LOD = 2.57) (Baghdadi *et al.,* 2006) and Chr20p13-12.3 (LOD = 3.33) (Mahasirimongkol *et al.,* 2009).

Whole-exome sequencing (WES) consists of capturing, sequencing and analysis of all exons (approx. 1%), a part of the human genome which is thought to be medically most applicable. This is thought to be the most applicable way of dealing with identification of genes that form a basis of diseases such as cancer, tuberculosis, and other rare diseases such as schizophrenia and autism, and many more. Exome sequencing study of pulmonary tuberculosis patients has beenperformed in Indian Bengali population which is homogenous, and are not known to be mixed by any other population. The study has identified polymorphisms in SIGLEC15 and HLA-DRA genes, which is understood to be novel in its association with tuberculosis. Haplotyping experiment was also performed which identified variants (rs13209234 and rs3177928) which discovered a significant relation with “AAGA” haplotype for the population (Bhattacharya *et al.,* 2019).

Next-generation sequencing (NGS) technologies have been extensively used to identify variants related to cancers. A search fornovel risk variants for Breast Cancer (BC) led to identification ofspecific variants associated with BC (Chandler *et al.,* 2015). Ng and colleagues did WES to studyMiller syndrome,a rare genetic condition, in four unrelated individuals and identified the causal variants (Ng *et al.,* 2009). Exome sequencing was performed on 250 patients who showed an impression of undiagnosed genetic conditions and achieved 25% molecular diagnostic success rate (Yang *et al.,* 2013). WES has a wide range of applications, it

is used in prenatal diagnosis as families and clinicians wish to uncover the genetic birth defectsto look for an explanation for the observed defects (Drury *et al.,* 2015).

The complete human mitochondrial DNA (mtDNA) wasfirst sequenced in 1981 (Sosa *et al.,* 2012). It is a circular genome, which is 16.5 kb in size. MtDNA is divided intonon-coding displacement loop commonly known as D-loop, constituting only 6.8% of the entire mt genome. This region controls replication and transcription of mtDNA. The coding region is larger than the non-coding region, constituting about 93.2% of the genome. The coding regionhousesthirty-seven (37) genes which is further of different forms such as 22 tRNAs, 13 proteins and 2 rRNAs that encode proteins whose expression leads to condition adverse to the electron transport chain. During oxidative phosphorylation, the mtDNAs are in continuous exposure to reactive oxygen species, which makes them highly susceptible to mutations (Liu *et al.,* 2001). It has been suggested that a higher fraction of mitochondrial mutation is likely to be functionally damaging and have distinct phenotypes as compared to nuclear mutations (Sui *et al.,* 2006).Thus, serving as an outstanding 'model system' for identification of genetic changes that will help solve the medical problems that genomic medicine will encounter.

Studies have been carried out to see the relation of mtDNA variations with different diseases, using deep sequencing. Study of Oral cancer showed that an increased in lymph node metastasis was contributed by non-synonymous mutations in mitochondrial respiratory genes (Palodhi *et al.,* 2018). Mitochondrial DNA mutations in gastric cancersuggests that the mitochondrial genomecan be used as a biomarker for early diagnosis of cancer. This is because at the earliest stages of cancer, the mt genome is susceptible (Sui *et al.,* 2006). In ovarian cancer, the study of mutations in mtDNA revealed somatic mtDNA mutations in 6 out of 10

tumorsamples, suggesting that in ovarian cancer, mtDNA mutation may be represented in 12s and 16s rRNA genes, the D-loop and the cytochrome b gene. Instability of mtDNA might leadtotumorigenesis (Liu *et al.,* 2001). Not much work has been carried out for tuberculosis, to identify the underlying genetic variants in mitochondrial genome in any population across the world.

Next-generation sequencing technology is being used in the clinical setting for specific medical conditions, such as identification of mutations contributing to rare genetic diseases. Promising results from the systematic use of NGS in patients, is already reported in countries such as USA, France and the Netherlands. The advancement in technologies is providing sufficiently great opportunities to execute personalized health strategies, prevention, early diagnosis, improved health maintenance, and development of therapy for patients with genetic diseases. Understanding the pathogenesisof tuberculosis and development of a plan of actionto achieve a long-term or overall aim to stop the disease burden and a better medical care given to a patient for the disease may be achieved by understanding the mechanism underlying genetic variations.

There are no scientific reports or evidence for host genetic factors linked to tuberculosis in Mizo population and this is the first study to explore the possible genetic variants which could predispose the population.

Aim and Objectives

The primary aim and objective of this study was:

- Identification of TB associated host candidate gene variations and their epidemiological interactions in Mizo Population.

Materials and Methods

Specimen collection and storage

Present study included all tuberculosis patients from different parts of Mizoram, coming to visit TB-DR centre, Falkawn, Mizoram. The age range of the patients were 15 – 85 years with a mean age of 43 years. The blood samples were collected in 2ml EDTA vials and immediately stored at -20 degrees until further processing. A questionnaire form along with written consent was filled up by each participant. Information regarding their personal details, health status like HIV, diabetes and BCG vaccine uptake, habits like tobacco chewing, smoking and alcohol consumption, diet, present complaint, history of past illness and family details was recorded.

Ethical statement

Approval for the study was obtained from the Institutional Ethics Committee (IEC) of Civil Hospital, Aizawl, Mizoram (B.12018/1/13-CH (A)/IEC/63, Dated: 28-03-2017). Blood was collected with written informed consent from each study participants.

DNA isolation

DNA isolation from blood was done using QIAamp (Qiagen) DNA Blood isolation kit (Cat No: 51106) following the protocol as manufacturer's recommendations. The blood samples were thawed at 4^0C and water bath was kept ready at 56^0C. QIAGEN protease, buffer AW1 and AW2 were prepared according to

instruction. 20µl of QIAGEN protease was pipetted in a 1.5ml Microcentrifuge tube (MCT) and 200µl of whole blood sample was added to it and mixed by pulse vortexing. After which, 200µl of AL buffer was added and pulse vortexed for 15 seconds for proper mixing, followed by incubation in water bath for 30 minutes at 56℃. After 30 minutes, the tubes were pulse spinned for maximum 30 seconds and 200µl of ethanol (chilled) was added and pulse vortexed. The entire lysate was then transferred into a QIAampcolumn placed in a 2ml collection tube. After closing the cap of spin column, the sample was centrifuged at 6000xg for 1 minute at Room Temperature (RT). After discarding the 2ml collection tube, the spin column was placed back in a new 2ml collection tube and 500µl of buffer AW1 was added into QIAamp column and centrifuged at 6000xg for 1 minute at RT. The 2ml collection tube was discarded and the column was placed back in a new 2ml collection tube. 500µl of buffer AW2 was added into the spin column and centrifuged at 20,000xg for 3 minute at RT. The flow through was discarded and returned the column back in the same collection tube.Centrifugation was done at 20,000xg for 1minute at RT. The 2ml collection tube was again discarded and the column was placed back in a new 1.5ml MCT, and dried for 15 minutes. 200µl of buffer AE was added directly into the membrane of the column and incubated at RT for 5 minutes and centrifuged at 6000xg for 1 minute at RT and the eluted genomic DNA was stored at - 20℃.

Statistical analysis

For performing statistical analysis IBM SPSS Statistics, version 20.0 (IBM Corp., Armonk, NY, USA) was used. The analysis was done for 274 cases and 500 controls. Odds ratios, 95% confidence intervals (95% CI) and *P* values were computed by 2- tailed Logistic Regression (Tables 2and 3). The validity of the

statistical model was evaluated through receiver operating characteristic curve (ROC curve) and the area under the ROC curve (AUC) was measured (Figure1).

Whole Exome Sequencing

A total of 10 TB patient's and 2 healthy control blood samples were recruited for this study. TruSeqExome Library Prep Kit (Catalog # FC-150-1004, Illumina) was used for library preparation. Assessment of quality and concentration of the DNA samples were performed using absorbance at 260 nm and 280 nm by Nanodrop Spectrophotometer (Nanodrop) and by Qubit HS kit. For each sample, 100 ng of DNA in 50µl shearing buffer was used for fragmentation of genomic DNA by ultra-sonication method using Covaris, followed by purification using Ampure XP (Beckman Coulter). End Repair, Adenylation, Adapter Ligation and Enrichment of adapter ligated DNA were performed. The validation of the genomic DNA library was performed by measuring the quantity of DNA library by Qubit and quality was verified in Bioanalyzer using High Sensitivity Chip (Agilent). After this, the pooling of adapter ligated DNA was completed according to the procedure mentioned in the above-mentioned guide. The pooled amplified DNA underwent two rounds of hybridization followed by PCR amplification and purification. Quantitation and evaluation of quality of the pooled captured library was performed with Qubit HS kit (Thermo) and Bioanalyzer using High Sensitivity Chip (Agilent). Sequencing (2 x 100 bp) of the pooled captured libraries was performed in HiSeq-2500 (Illumina). The sequencing was performed at National Institute of Biomedical Genomics (NIBMG), Kalyani, West Bengal. The whole exome sequence data analysis involved pre-processing of the sequence file. For the preliminary quality check of FASTQ files that has been generated from Illumina HiSeq-2500, FASTQC was used (Andrews, 2010), and the low quality reads was removed by Trimmomatic software (AVG =

Q30, MINLENGHT = 50). Processed fastq files were mapped using BWA on human reference sequence (hg19). Coversion of SAM file to BAM files was done usingSAMtools (Li *et al.,*2009). BAM files were quality checked using QualiMap to ensure good sequence coverage with the reference sequence (García-Alcalde *et al.,* 2012), followed by variant calling using GATK. Annotation of the variants was performed with Annovar (Wang *et al.,* 2010). LJB database was used to identify the previously reported variants and also to predict the effect of these variants on the gene function (http://annovar.openbioinformatics.org/en/latest/user-guide/filter/#ljb42-dbnsfp-non-synonymous-variants-annotation). Further analysis of the variants was done through ClinVar for studying the relationships between clinically relevant variants and phenotypes (Landrum *et al.,* 2014).

Filtering approaches

The data was filtered using different approaches to help prioritize variants. The first filter (Filter1) involved removal of all the common variants between patient's and healthy controls. Such filtering helps in reducing the variants list, giving only case specific variants. The second filter (Filter2) was done using HGV & TB database: a online resource for human genes and gene variants associated with tuberculosis (Sahajpal *et al.,* 2014). This gives a list of genes reported to be associated with tuberculosis. The third filter (Filter3) eliminated the variants which are not reported to be pathogenic, thus giving only pathogenic variants reported across the populations (Table 4).

To filter out unreported variants, again different filtering approaches were used. The first filter (filter1) was same as above mentioned filter1. Then, the next filter (Filter2) for unreported variants, involved removal of variants which has no Reference SNP cluster ID (rsid) which were thought to be unreported. Variants were

further segregated (Filter3) based on population frequency obtained from databases such as refGene (http://varianttools.sourceforge.net/Annotation/RefGene), dbSNP (https://www.ncbi.nlm.nih.gov/snp), 1000 Genomes (Gibbs *et al.,* 2015), ExAC (http:// exac.broadinstitute.org),ESP6500(https://evs.gs.washington.edu/EVS/), GWAS (https://www.ebi.ac.uk/gwas/) and gnoMAD (https://gnomad. broadinstitute.org/) and also using function prediction tool such as SIFT(https://sift.bii.a-star.edu.sg/), Polyphen2 (http://genetics.bwh.harvard. edupph2/), PROVEAN (http://provean.jcvi.org/index.php), Mutation Taster (http://www.mutationtaster.org/), Fathmm MKL(fathmm.biocompute. org.uk/fathmm MKL.htm), CADD (https://cadd.gs.washington.edu) and DANN (Quang *et al.,* 2015) (Table 5).

The genes were subjected to VarElect: Next Generation Sequencing Phenotyper to prioritize the gene variants based on the disease phenotype (http://varelect.genecards.org/). Some of the unreported variants were analysed through HOPE software (http://www.cmbi.umcn.nl/hope) that predicts the structural and functional effects of the variants (Table6).

Whole Mitochondrial DNA Sequencing

A total of seven TB patient's and twelve healthy controls blood samples were recruited for the study. SequalPrep™ Long PCR Kit with dNTPs (Thermo Scientific) was used for amplification of whole mtDNA (16.6kb) from blood in ABI 9700 thermal cycler (Thermo Scientific). Visualization and purification of the PCR products were done in 0.8% agarose gel electrophoresis and QIAquick Gel Extraction Kit (Qiagen), respectively. The purified amplicons were processed for sequencing library preparation using Nextera XT DNA library kit (Illumina). After purification, the libraries were pooled and sequenced in IlluminaHiSeq – 2500 to generate 2 x 100

bp reads. The sequencing was performed at National Institute of Biomedical Genomics (NIBMG), Kalyani, West Bengal.

The preliminary quality checkingof the FASTQ files generated from Illumina HiSeq-2500 was done using FASTQC (Andrews, 2010).BWAMEM tool was used for alignment of the sequence data to the mitochondrial reference sequence (rCRS-revised Cambridge Reference Sequence) (Li *et al.,* 2013). SAMtoolswas used to convert SAM file to BAM files, and reads below mapping quality of 40 were discarded (Li *et al.,* 2009). QualiMapwas used to check the quality of BAM filesto align with the mtDNA reference sequence (García-Alcalde*et al.,* 2012). Sequence data in BAM format has been deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), hosted by the EBI and the accession number is awaited. VarScan2 was used for detection of variants in the mtDNA (Koboldt*et al.,* 2012). Integrative Genome Viewer (Robinson *et al.,* 2011) and ANNOVAR (Wang *et al.,* 2010) was used for the visualization and annotation of variants respectively.

Filtering approaches

The mitochondrial sequence data was filtered using different approaches to help prioritize variants. The first filter (Filter1) involves removal of all the common variants between patients and healthy controls. Such filtering helps in reducing the variants list giving only case specific variants (Tables7 and 8). After filter1, non-synonymous variants underwent filtering through function prediction tools such as Polyphen2 (http://genetics.bwh.harvard.edu/pph2/) and PROVEAN (http://provean.jcvi.org/index.php) (Table 9).

The non-synonymous variants, which were deleterious were analysed through HOPE software (http://www.cmbi.umcn.nl/hope)that predicts the structural and functional effects of the variants (Table10).

Results

Demography and lifestyle status of patients

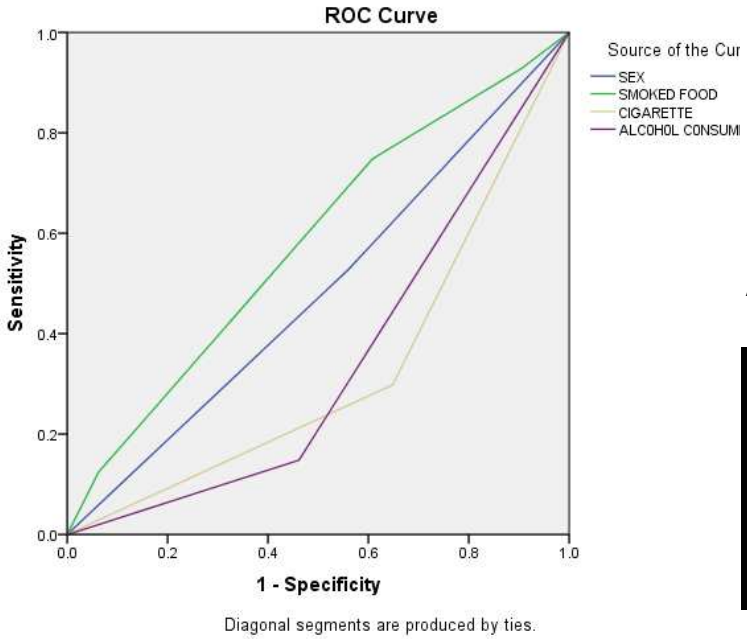
Out of 274 cases, 153 were males and 121 females that were involved for the study.Most of the study samples were smokers (58.8%), alcohol users (54.0%), and other tobacco product users (38.7%). Most of the patients consumed fermented foods (50.7%) and smoked foods (54.4%) for about 2-4 days in a week. The statistical analysis showed that Sex (Male), Alcohol consumption, Smoking and consumption of smoked foods were statistically significant and might be therisk factors for the disease. Detailed information on demography and lifestyle status analysis is given in Tables 2 and 3. The mathematical calculation was validated using ROC curve (Figure1).

Table 2: Univariate analysis to establish a mathematical model between an independent variable and TB.

		Cases (274)	Controls (500)	P-value	ORs (95% CI)
Sex	Female **Male**	121(44.2%) **153(55.8%)**	264(52.8%) **236(47.2%)**	**0.001**	Reference **0.425 (0.272 – 0.663)**
Alcohol	No **Yes**	121 (40.5%) **148 (54.0%)**	74 (14.8%) **426 (85.2%)**	**0.001**	Reference **5.293(3.278 – 8.546)**
Smoking	No **Yes**	83 (30.3%) **161(58.8%)**	149 (29.8%) **351 (70.2%)**	**0.001**	Reference **4.170 (2.713 – 6.411)**
Other tobacco products	No **Yes**	52 (19.0%) **106 (38.7%)**	152 (30.4%) **348 (69.6%)**	**0.011**	Reference **0.602(0.406-0.891)**
Tuibur	No **Yes**	56(20.4%) **33(12.0%)**	384(76.8%) **116(23.2%)**	**0.003**	Reference **0.448 (0.266 – 0.754)**
Meat consumption	Never 1 day in a week 2-4 days in a week 5-7 days in a week	3 (1.1%) 96 (35.0%) 139 (50.7%) 35 (12.8%)	101 (20.2%) 184 (36.6%) 215 (43.0%) 215 (43.0%)	0.649	Reference 0.948 (0.753 – 1.194)
Fermented foods	Never 1 day in a week **2-4 days in a week** 5-7 days in a week	29 (10.6%) 116 (42.3%) **122 (44.5%)** 6 (2.2%)	68 (13.6%) 367 (73.4%) **63 (12.6%)** 2 (4%)	**0.001**	Reference **3.704 (2.648 – 5.180)**
Smoked foods	Never 1 day in a week **2-4 days in a week** 5-7 days in a week	26 (9.5%) 81 (29.6%) **149 (54.4%)** 17 (6.2%)	36 (7.2%) 90 (18.0%) **312 (62.4%)** 62 (12.4%)	**0.001**	Reference **0.492 (0.371 – 0.653)**

Table 3: Multivariate analysisto establish a relationship among the major risk factors involved in TB infection

		Cases (274)	Controls (500)	P-value	ORs (95% CI)
Sex	Female	121(44.2%)	264(52.8%)		Reference
	Male	**153(55.8%)**	**236(47.2%)**	**0.001**	**2.240 (1.420 – 3.534)**
Alcohol	No	121 (40.5%)	74 (14.8%)		Reference
	Yes	**148 (54.0%)**	**426 (85.2%)**	**0.001**	**5.293(3.278 – 8.546)**
Smoking (Cigarette)	No	83 (30.3%)	149 (29.8%)		Reference
	Yes	**161(58.8%)**	**351 (70.2%)**	**0.001**	**0.254 (0.169 – 0.381)**
Other tobacco products	No	52 (19.0%)	152 (30.4%)		Reference
	Yes	106 (38.7%)	348 (69.6%)	0.600	0.353 (0.007 – 1.194)
Tuibur	No	56(20.4%)	384(76.8%)		Reference
	Yes	33(12.0%)	116(23.2%)	0.003	0.448 (0.266 – 0.754)
Fermented foods	Never	29 (10.6%)	68 (13.6%)		Reference
	1 day in a week	116 (42.3%)	367 (73.4%)		
	2-4 days in a week	122 (44.5%)	63 (12.6%)	0.409	0.425 (0.056 – 3.250)
	5-7 days in a week	6 (2.2%)	2 (4%)		
Smoked foods	Never	26 (9.5%)	36 (7.2%)		Reference
	1 day in a week	81 (29.6%)	90 (18.0%)		
	2-4 days in a week	**149 (54.4%)**	**312 (62.4%)**	**0.001**	**0.492 (0.371 – 0.653)**
	5-7 days in a week	17 (6.2%)	62 (12.4%)		



Area Under the Curve

Test Result Variable(s)	Area
SEX	0.484
SMOKED FOOD	0.582
CIGARETTE	0.325
ALC0H0L C0NSUMPTI0N	0.343

Figure 1: Receiver operating characteristic (ROC) curve of the evaluation model.

Whole exome sequencing

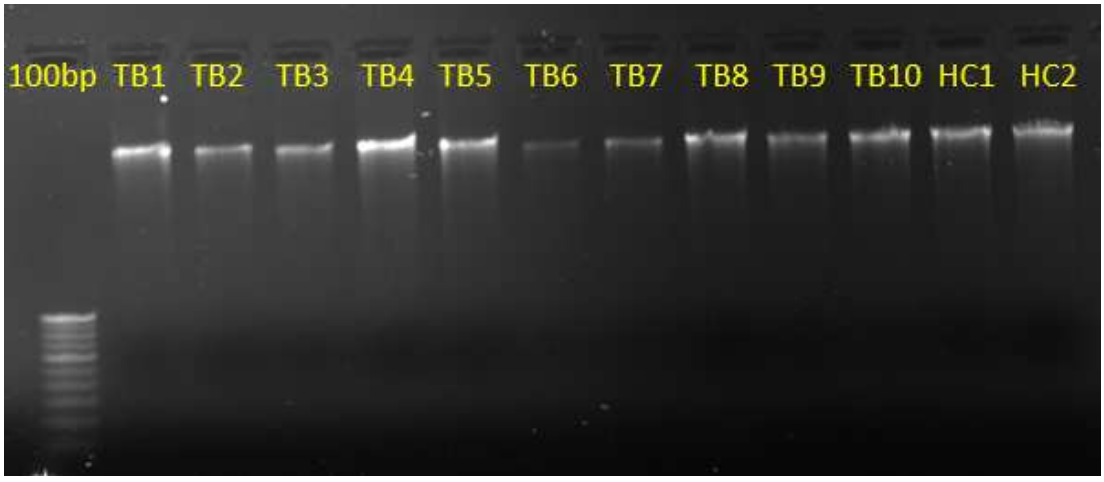


Figure 2: Agarose gel (0.8 %) showing the presence of genomic DNA isolated from patient's blood samples and healthy control.

Upon Whole exome sequencing, about 30,000variants were detected per sample. The first filtering approach (Filter1) reduced the list to 7,000-8,000 variants per individual. After applying filter2, all the known variants which were known to have an association withTB across different populations were given and further analysis using filter3 gave all the reported pathogenic variants across the samples (Table 4). All the patients were found to harbour, P2RX7 and VDR gene variants reported to be pathogenic. SFTPA1 gene variant was found in 80% of the patient, WIPI11 gene variant in 70% of the patient, SP110 gene variant in 60% of the patient, MC3R, P2RX7 and SLC11A1 gene variants was found in 40% of the patient, MBL2 and TIRAP gene variant in 30% of the patient, MACRO and SFTPA2 gene variant in 20% of the patient and LAMP1,IL1B,IL10RA and SFTPA1 in 10% of the sample.

For unreported variants, after obtaining 7,000-8,000 variants from filter1, these variants were segregated using filter2 which gave about 250-320 variants which did not have any rsid. Filter3 gave about 10-30 variants which shows to be probably damaging/pathogenic/deleterious depending on the prediction tools used and also considering the population frequency being “0” in the databases that has been used.

The detailed information for each individual is given in table 5.

Table 4: List of pathogenic variants as reported by HGV&TB database

Gene(s)	Ref	Alt	Amino Acid Change	TB1	TB2	TB3	TB4	TB5	TB6	TB7	TB8	TB9	TB10
MC3R	G	A	V44I	√	√	-	√	-	-	-	√	-	-
P2RX7	A	T	N568I	√	√	√	√	√	√	√	√	√	√
LAMP1	C	T	A204V	√	-	-	-	-	-	-	-	-	-
MBL2	C	T	G54D	√	-	-	-	√	-	-	-	√	-
TIRAP	G	A	S55N	√	√	-	-	-	-	-	√	-	-
VDR	T	C	M1T	√	√	√	√	√	√	√	√	√	√
SFTPA1	C	T	R219W	-	√	√	√	√	√	√	**-**	√	√
SP110	A	G	M523T	-	√	√	-	-	-	√	√	√	√
WIPI11	G	A	T31I	-	√	**-**	√	√	√	√	√	√	**-**
MACRO	T	C	F282S	-	√	-	-	-	√	-	-	-	-
P2RX7	A	C	M163V	-	-	√	-	-	√	-	√	√	-
SLC11A1	G	A	D543N	-	-	√	-	-	√	-	√	√	-
IL1B	G	T	G31V	-	-	-	-	√	-	-	-	-	-
IL10RA	A	G	S159G	-	-	-	-	-	-	√	-	-	-
SFTPA1	C	G	L50V	-	-	-	-	-	-	-	√	-	-
SFTPA2	G	C	A91P	**-**	-	-	-	--	-	-	-	√	√

TB1 to TB10 –Patient samples

Ref - Reference Allele

Alt- Altered Allele

Table 5: List of variants analysed to be pathogenic and not reported for any population.

Sample ID	Gene	Position	Reference	Alternate	Amino Acid Change
TB-1	SPEN	16260384	C	T	T2550I
	TRIM46	155150484	G	T	G180C
	MIA3	222827795	C	G	S1481C
	BEND4	42154127	G	A	P12S
	ELL2	95249510	C	A	S149I
	PCDHB2	140476318	C	A	D648E
	DXO	31937895	T	C	D345G
	PACRG	163736000	A	G	E291G
	DNAJC2	102953480	A	G	Y569H
	ADAM28	24177809	G	C	D213H
	BSPRY	116111984	C	A	A42D
	SPAG6	22680773	T	G	V374G
	FRAT1	99079869	T	G	L220R
	PCDH8	53421034	C	T	G513D
	TGM5	43525789	C	T	E658K
	MID2	107169912	G	T	G576V
TB-2	RNF207	6272785	T	C	I509T
	HES6	239148474	C	G	D25H
	LSAMP	115560820	A	T	L264H
	PPIC	122365004	C	G	G76A
	PCDHA5	140203252	G	A	S631N
	CHD7	61778233	G	T	G2912V
	ALG10B	38714423	G	A	G277D
	CSRNP2	51461469	C	A	G232V
	OR6C4	55945540	G	C	C177S
	TMCC3	94976080	T	C	K74E
	METTL22	8729148	G	A	A227T
	HOXB13	46805483	G	C	P158R
	HEATR6	58134758	T	A	N577I
	FAM69C	72114499	C	A	G73V
TB-3	TARS2	150471112	A	G	D328G
	SLC27A3	153748276	C	A	N148K
	ZCCHC4	25315743	G	C	R68T
	RRH	110754371	T	A	N61K
	NKAPL	28227780	G	T	D211Y
	RARS2	88224213	C	T	R552K
	FBXO43	101146142	A	T	L672Q
	CEP78	80880328	G	T	D557Y
	PDDC1	774070	C	T	R62H
	ARHGAP20	110450870	T	G	S911R
	BLVRB	40953923	A	C	V165G
	PMEPAI	56227195	G	A	R232W

TB-4	MMEL1	2525853	G	A	A573V
	TITHD1	24105019	A	G	H5R
	COA7	53153667	C	G	A141P
	FAM69A	93341893	T	C	Y50C
	PCDHGA7	140764128	C	A	D554E
	NDUFA5	123185678	T	C	E64G
	NOV	120435166	A	T	R290W
	NOV	120435167	G	T	R290M
	CEP78	80880328	G	T	D557Y
	BTAF1	93726465	G	A	R534Q
	TCIRG1	67817661	G	C	G510R
	MNAT1	61346454	C	A	P195H
	BEGAIN	101012907	C	A	R36L
	DNASE1	3707935	T	C	V277A
	STAT5B	40370768	G	A	T321M
	MFSD11	74740433	C	G	L124V
	ABCA7	1041884	G	T	G72V
	EMR2	14862452	T	G	H549P
TB-5	KLHL17	899899	G	T	M563I
	UGP2	64114711	T	C	L405P
	DNAH11	21775426	C	A	P2537T
	DNAH11	21784534	A	G	H2788R
	PLEC	144995785	G	T	A2721D
	GLIPR2	36148647	G	A	G76R
	TNKS2	93558643	G	A	A66T
	MNAT1	61346454	C	A	P195H
	GLCE	69561231	C	A	P501Q
	SMG8	57290159	G	C	D659H
	TBC1D16	77918823	C	A	G223C
	USP14	210394	T	A	S377T
	KCNS1	43723830	G	A	T421I
TB-6	FAM69A	93341893	T	C	Y50C
	NCOA1	24881617	G	A	C24Y
	VPRBP	51457569	G	A	P898L
	TAGAP	159462546	T	C	D43G
	LRRC6	133622471	C	T	D361N
	CTNNA3	68526119	G	C	P395R
	MRVI1	10631302	A	G	L200P
	WDR89	64066402	A	G	S87P
	BTBD7	93717962	G	A	L246F
	EHD4	42246064	A	G	M104T
	FAM69C	72114499	C	A	G73V
	CAMSAP3	7670130	T	C	L56P
TB-7	CNKSR1	26515401	A	G	N167S
	OBSCN	228559093	A	C	S6872R
	BIRC6	32640590	A	G	H744R
	EML4	42508012	T	A	F172L

TB-7	BAZ2B	160295044	G	C	P353A
	KLHDC8B	49213076	A	T	S303C
	SLC15A2	121641665	C	T	P244L
	PARP9	122247485	G	T	A764D
	PDIA5	122880764	A	T	E506V
	PDLIM3	186423510	C	G	A178P
	RANBP17	170345843	T	A	F361I
	IQUB	123136774	C	T	E404K
	SVOPL	138312929	A	T	I196N
	FBXW5	139836830	G	A	A255V
	LYZL2	30915774	C	G	R70P
	LIPK	90503060	C	A	N314K
	METTL1	58163391	C	A	Q115H
	OSBPL8	76791619	A	G	L134P
	LTBP2	75052806	G	C	P194R
	SLC28A1	85478386	G	T	W448C
	DNASE1	3707935	T	C	V277A
	ITGAD	31418984	A	T	I285F
	ALDH3A2	19566658	T	C	L318P
	RAD51D	33428357	C	G	D144H
	KRTAP1	39191016	A	T	C20S
	FAM171A2	42431955	G	T	P543T
	LONP1	5692200	G	A	T712M
	VMAC	5905088	G	T	D63Y
	DEPDC5	32241165	G	A	R910H
TB-8	TTN	179455053	C	T	A11402T
	OBSL1	220431725	A	C	L654R
	CHST13	126260841	T	C	L149P
	FAM160A1	152559891	G	A	M403I
	MAN1A1	119669645	T	C	R196G
	SH3GL2	17795679	C	T	H333Y
	PIK3C2A	17139076	C	T	A1060T
	CCDC91	28412325	G	A	G20D
	IRS2	110436060	G	C	P781A
	WDR89	64066367	A	T	D98E
	TLN2	63058601	G	A	A1726T
	MFSD11	74740433	C	G	L124V
	LOXHD1	44057649	G	A	T442I
	SALL3	76755388	C	A	R1133S
TB-9	PRG4	186280235	G	T	G1056V
	CDH10	24498624	A	T	N466K
	NIM1K	43277180	T	C	L105P
	STYXL1	75630254	G	T	A255D
	FBXL13	102524759	T	A	I337F
	KIF27	86506269	G	A	P584S
	SOX6	16010574	G	C	D618E
	UGGT2	96536758	A	G	I1072T

	EHD4	42264519	G	C	F58L
	MTHFSD	86585826	G	T	P66T
	RAD51D	33428357	C	G	D144H
	INPP5J	31523949	G	C	W165C
	GGA1	38028615	T	C	V519A
TB-10	ANKRD13C	70819734	C	A	V120F
	CCDC74B	130897841	G	C	P197A
	DVL3	183884446	T	C	F345L
	GPX8	54456947	G	C	Q110H
	P4HA2	131539838	C	T	R363Q
	PCDHGA7	140764128	C	A	D554E
	MICALL2	1477605	C	T	D775N
	MEST	130139728	A	G	E169G
	KRT18	53343084	G	T	G43C
	PTPN21	88946079	G	A	P566S
	CHRNA7	32393511	G	T	E67D
	CSPG4	75967949	A	G	L2304P
	C15orf40	83677379	T	A	E96V
	LYSMD4	100272199	C	A	R2S
	EPS15L1	16466540	G	T	S903R
	ZNF675	23836411	G	C	H442D
	GRIK5	42563679	C	A	C170F
	SCRT2	644370	G	C	A290G
	GPC4	132458345	T	C	Y180C

TB1 to TB10 – Patient samples

Table 6: The list of gene variants and its structural prediction using HOPE software.

Gene	Amino Acid change	Domain	Function
PACRG	E291G	Parkin Co-Regulated Protein IPR019399	The mutation introduces a glycine at this position. Glycines are very flexible and can disturb the required rigidity of the protein at this position. The charge of the wild-type residue will be lost, this can cause loss of interactions with other molecules or residues. The mutant residue is smaller, this might lead to loss of interactions. The hydrophobicity of the wild-type and mutant residue differs. The mutation introduces a more hydrophobic residue at this position. This can result in loss of hydrogen bonds and/or disturb correct folding.
TCIRG1	G510R	V-Type Atpase, V0 Complex, 116Kda Subunit Family IPR002490	The mutant residue is bigger than the wild-type residue. The mutant residue introduces a charge in a buried residue which can lead to protein folding problems. This means that in some rare cases the mutation might occur without damaging the protein. The wild-type residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function.
TRIM46	G283C		The wild-type residue is a glycine, the most flexible of all residues. This flexibility might be necessary for the protein's function. Mutation of this glycine can abolish this function. The mutant residue is bigger, this might lead to bumps. The torsion angles for this residue are unusual. only glycine is flexible enough to make these torsion angles, mutation into another residue will force the local backbone into an incorrect conformation and will disturb the local structure.

SLC27A3	F148K	Solute Carrier Family 27 Member 3 IPR030307	The wild-type residue is predicted (using the Reprof software) to be located in its preferred secondary structure, a β-strand There is a difference in charge between the wild-type and mutant amino acid. The mutation introduces a charge, this can cause repulsion of ligands or other residues with the same charge. The hydrophobicity of the wild-type and mutant residue differs. Hydrophobic interactions, either in the core of the protein or on the surface, will be lost.

Whole mitochondrial sequencing

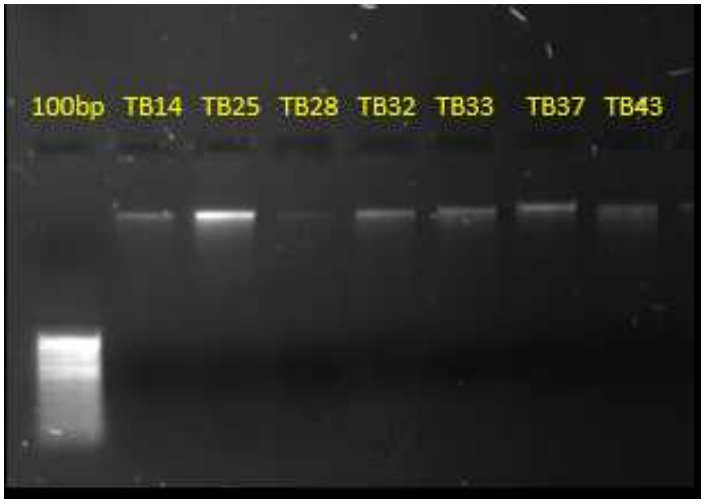


Figure 3: Agarose gel (0.8 %) showing the presence of genomic DNA isolated from patient's blood samples (TB).

Upon mtDNA sequencing filtering of variants resulted in non-coding (Table 6) and coding mutations (Table 7). The non-synonymous mutations filtered through prediction tools are listed in Table 8. In the non-coding region, total of sixty-five variants were present in the D-loop region, fourteen variants in 16s, three in 12s, two in tRNA-Thr and tRNA-Gln. And for coding regions, a total of ninety-one variants were present in different genes, out of which sixty-four were non-synonymous and twenty-seven were synonymous. Out of twenty-seven non-synonymous mutations: MT-ND2 g. A>G4824 p.T119A, found in two patients (TB-25 and TB-37), and MT-ND6 g. T>C14180 p.Y165C, found in one patient (TB-25) was predicted **deleterious** in Polyphen-2 and Provean.

Table 7: Mutations in the non-coding region of the mtDNA in TB patients.

Patient ID	Position	Reference	Variant	Gene	VarFreq
TB-14	152	T	C	D-loop	99.89%
	194	C	T		99.8%
	247	G	-A		91.05%
	279	T	C		99.87%
	489	T	C		99.94%
	2706	A	G	16s	99.96%
	15926	C	T	tRNA-Thr	99.89%
	16223	C	T	D-loop	99.59%
	16234	C	T		99.92%
	16325	T	C		99.86%
	16390	G	A		99.81%
	16519	T	C		99.88%
TB-25	302	A	AC		8%
	385	A	G		99.81%
	489	T	C		99.93%
	2150	T	TAA	16s	91.36%
	2706	A	G		99.91%
	3105	A	-CC		99.02%
	16183	A	AC	D-loop	8.68%
	16189	T	C		99.54%
	16215	A	G		99.54%
	16223	C	T		99.52%
	16316	A	G		99.92%
	16325	T	C		99.9%

TB-28	302	A	AC		8.03%
	663	A	G	12s	99.67%
	2706	A	G	16s	99.96%
	16179	C	-A	D-loop	16.06%
	16172	T	C	D-loop	99.51%
	16356	T	C	D-loop	99.84%
	16362	T	C		99.84%
	16519	T	C		99.83%
TB-32	152	T	C		99.81%
	247	G	-A		95.11%
	302	A	AC		9.31%
	489	T	C		99.91%
	2352	T	C	16s	99.91%
	2706	A	G		99.95%
	2780	C	T		99.8%
	4363	T	C	tRNA-Gln	91.99%
	16185	T	C	D-loop	97.55%
	16189	T	C		99.63%
	16223	C	T		99.7%
	16260	C	T		99.8%
	16298	T	C		99.82%
TB-33	152	T	C		99.91%
	234	A	G		99.97%
	247	G	-A		90.5%
	302	A	AC		9.06%
	513	G	CA		84.76%
	745	A	AT	12s	99.89%
	1927	G	A	16s	99.91%
	2706	A	G	16s	99.98%
	10454	T	C	D-loop	99.92%
	16093	T	C		99.29%
	16111	C	T		97.98%
	16129	G	A		99.61%
	16304	T	C		99.68%
	16311	T	C		99.88%
	16519	T	C		99.82%
TB-37	152	T	C		99.7%
	200	A	G		97.83%
	235	A	G		100%
	302	A	AC		6.27%
	513	G	-CA		84.6%
	663	A	G	12s	99.82%
	2706	A	G	16s	99.94%
	16223	C	T	D-loop	99.65%
	16290	C	T		99.77%
	16319	G	A		99.71%
	16362	T	C		99.98%

TB-43	152	T	C		99.92%
	247	G	-A		90.9%
	302	A	AC		9.41%
	489	T	C		99.91%
	2352	T	C	16s	99.93%
	2706	A	G		99.95%
	2780	C	T		99.84%
	4363	T	C	tRNA-Gln	99.88%
	16223	C	T	D-loop	99.75%
	16260	C	T		99.89%
	16298	T	C		99.88%
	16519	T	C		99.84%

Table8 : Mutations in the coding region of the mtDNA in TB patients.

Patient ID	Position	Reference	Variant	Gene	Type	VarFreq
TB-14	3780	C	T	ND1	Synonymous	99.87%
	6179	G	A	COX1	Synonymous	99.86%
	7853	G	A	COX2	Nonsynonymous	99.83%
	8701	A	G	ATP6		99.87%
	9469	C	T	COX3		99.81%
	9540	T	C	COX3	Synonymous	99.92%
	10398	A	G	ND3	Nonsynonymous	99.92%
	10400	C	T	ND3	Synonymous	99.84%
	12705	C	T	ND5		99.91%
	13356	T	C	ND5		99.94%
	14233	A	G	ND6		99.93%
	14783	T	C	CYTB		99.94%
	15043	G	A			99.89%
	15301	G	A			99.93%
	15326	A	G		Nonsynonymous	99.93%
TB-25	3999	T	C	ND1	Synonymous	99.88%
	4907	T	C	ND2		99.92%
	5231	G	A	ND2		99.79%
	8701	A	G	ATP6	Nonsynonymous	99.95%
	10398	A	G	ND3	Nonsynonymous	99.89%
	10400	C	T	ND3	Synonymous	99.77%
	11176	G	A	ND4		99.79%
	12705	C	T	ND5		99.94%
	12876	C	T			99.87%
	12417	C	-A		Frameshift substitution	3.52%
	14180	T	C	ND6	Nonsynonymous	99.93%
	14783	T	C	CYTB	Synonymous	99.97%
	15043	G	A			99.87%
	15301	G	A			99.8%
	15440	T	C			99.92%

	15530	T	C			99.88%
	15671	A	G		Nonsynonymous	99.95%
TB-28	4248	T	C	ND1	Synonymous	99.93%
	4824	A	G	ND2	Nonsynonymous	99.86%
	8794	C	T	ATP6	Nonsynonymous	99.9%
	10646	G	A	ND4L	Synonymous	99.88%
	11611	G	A	ND4		99.83%
	12417	C	-A	ND5		3.24%
	13708	G	A	ND5	Nonsynonymous	99.77%
	14569	G	A	ND6	Synonymous	99.8%
	15613	A	G	CYTB		99.85%
TB-32	4715	A	G	ND2		99.82%
	4841	G	A	ND2		99.93%
	6752	A	G	COX1		99.94%
	7196	C	A	COX1		99.86%
	8584	G	A	ATP6	Nonsynonymous	99.83%
	8598	T	C		Synonymous	99.92%
	8701	A	G		Nonsynonymous	99.93%
	8945	T	C		Nonsynonymous	99.93%
TB-32	9090	T	C	ATP6	Synonymous	99.93%
	9540	T	C	COX3	Synonymous	99.98%
	10398	A	G	ND3	Nonsynonymous	99.95%
	10400	C	T	ND3	Synonymous	99.81%
	10490	T	C	ND4L	Synonymous	99.93%
	10653	G	A	ND4L	Nonsynonymous	99.91%
	12417	C	-A	ND5	Synonymous	3.34%
	12705	C	T	ND5		99.85%
	14783	T	C	CYTB		99.95%
	15043	G	A			99.94%
	15301	G	A			99.9%
	15487	A	T			99.56%
	15784	T	C			99.95%
TB-33	3970	C	T	ND1		99.81%
	6392	T	C	COX1		99.92%
	6599	A	G			99.89%
	6962	G	A			99.89%
	9053	G	A	ATP6	Nonsynonymous	99.77%
	10609	T	C	ND4L	Nonsynonymous	99.94%
	12406	G	A	ND5	Nonsynonymous	99.97%
	12882	C	T		Synonymous	99.89%
	13759	G	A		Nonsynonymous	99.56%
	13928	G	C		Nonsynonymous	99.93%
TB-37	4248	T	C	ND1	Synonymous	99.87%
	4824	A	G	ND2	Nonsynonymous	99.86%
	8794	C	T	ATP6	Nonsynonymous	99.84%
	12705	C	T	ND5	Synonymous	99.97%
	15478	A	G	CYTB		99.92%
TB-43	4715	A	G	ND2		99.91%
	4841	G	A	ND2		99.89%
	7196	C	A	COX1		99.78%
	8584	G	A	ATP6	Nonsynonymous	99.76%
	8701	A	G		Nonsynonymous	99.98%

TB-43	8598	T	C		Synonymous	99.93%
	9090	T	C		Synonymous	99.88%
	9540	T	C	COX3	Synonymous	99.91%
	10398	A	G	ND3	Nonsynonymous	99.95%
	10400	C	T	ND3	Synonymous	99.84%
	10490	T	C	ND4L	Synonymous	99.93%
	10653	G	A	ND4L	Nonsynonymous	99.85%
	12705	C	T	ND5	Synonymous	99.87%
	14783	T	C	CYTB	Synonymous	99.95%
	15043	G	A			99.88%
	15301	G	A			99.93%
	15487	A	T			99.78%
	15784	T	C			99.88%

Table9 : Nonsynonymous mutations with their functional predictions in the mtDNA

Patient ID	Position	Ref	Alt	Gene	AA Change	Polyphen-2	Provean
TB-14	7853	G	A	COX2	V90I	Benign	Neutral
	8701	A	G	ATP6	T59A		
	9469	C	T	COX3	T88I		
	10398	A	G	ND3	T114A		
	15326	A	G	CYTB	T194A		
TB-25	8701	A	G	ATP6	T59A		
	10398	A	G	ND3	T114A		
	14180	**T**	**C**	**ND6**	**Y165C**	**Deleterious**	**Deleterious**
	15671	A	G	CYTB	M309V	Benign	Neutral
TB-28	**4824**	**A**	**G**	**ND2**	**T119A**	**Deleterious**	**Deleterious**
	8794	C	T	ATP6	H90Y	Benign	Neutral
	13708	G	A	ND5	A458T		
TB-32	8701	A	G	ATP6	T59A		
	8584	G	A	ATP6	A20T		
	8945	T	C	ATP6	M140T		
	10398	A	G	ND3	T114A		
TB-33	9053	G	A	ATP6	S176N		
	10609	T	C	ND4L	M47T		
	12406	G	A	ND5	V24I		
	13759	G	A	ND5	A475T		
	13928	G	C	ND5	S531T		
TB-37	**4824**	**A**	**G**	**ND2**	**T119A**	**Deleterious**	**Deleterious**
	8794	C	T	ATP6	H90Y	Benign	Neutral
TB-43	8584	G	A	ATP6	A20T	Deleterious	Deleterious
	8701	A	G	ATP6	T59A	Benign	Neutral
	10398	A	G	ND3	T114A		
	10653	G	A	ND4L	A62T		

Table 10: The list of gene variants and its structural prediction using HOPE software

Gene	Amino Acid change	Domain	Function
MT-ND2	T119A	Nadh:QuinoneOxidoreductase/MrpAntiporter, Membrane Subunit IPR001750	Mutant residue is located near a highly conserved position. The mutant residue is smaller than the wild-type residue. The mutation will cause an empty space in the core of the protein. The hydrophobicity of the wild-type and mutant residue differs. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.
MT-ND6	Y165C	Nadh:Ubiquinone/PlastoquinoneOxidoreductase, Chain 6 IPR001457	The mutant residue is smaller than the wild-type residue. This size difference can affect the contacts with the lipid-membrane. The mutant residue is more hydrophobic than the wild-type residue. The mutation will cause an empty space in the core of the protein. The mutation will cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding.

Discussion

In this study,a statistical relationship of the key known TB risk factors has been established between TB patients and controlsin Mizo population.Males constituted 55.8% of the patients in the study (Table 2). According to Global Tuberculosis Report 2018, adults are mostly infected by the disease contributing to about 90% of the cases, and there are more cases of than women. A meta-analysis in developed and developing countries across the world found that the rate of confirmed TB was 2.21 (1.92 - 2.54) timeshigher than adult women (Horton *et al.,* 2016). The present study showed that males are statistically more susceptible to TB as compared to females. There can be a number of reasons why this happens:the habit of smoking and drinking is common in Men and also because they are more socialized as compared to women and have a greater chance of being in close contact with the TB infected individuals (Dodd *et al.,* 2016). Difference in iron-metabolism among sex may also lead to higher TB burden in males. Iron is important for the growth of Mtb and as a result tries to access the iron store of the host. The hormone hepcidin is known to enhance the growth of the bacteria inside macrophages, thus the level of iron and hepcidin may predict the occurrence of the disease in a healthy individuals. The concentration of hepcidin varies with age and sex, and is known to be higher in men because of the absence of menstrual blood loss. Therefore, due to the higher level of iron the growth of bacteria is favourable in men and also because of men's larger lung volumes, thus making them more susceptible to the *Mtb* infection (Ganz *et al.,* 2015).

There is a higher risk of developing TB among people who have the habit of drinking alcohol, and/or have an alcohol use disorder. Various epidemiological

studies reported alcohol use as one of the important factor associated with tuberculosis. In the present study, more than 50% of the patients were alcohol users (Table2). Alcohol users are immuno-compromised and are particularly susceptible to infectious diseases such as TB as compared to the abstainers (Happel *et al.,* 2005; Zhang *et al.,* 2002). Areview work conducted by different researchers established that alcohol use was independently associated with TB (Lonnroth *et al.,* 2008; Rehm*et al.,* 2009). The role of alveolar macrophages are to destroy/ kill the inhaled *Mtb* bacteria, which multiplies in the lungs (Dannenberg *et al.,* 1989). Alcohol exposure supresses the capacity of the antigen-specific T-cells, protective effect of cytokines and other immune cells which has the ability to inhibit bacterial growth (Bermudez *et al.,* 1991; Szabo *et al.,* 1995; Neuman *et al.,* 2003; Crews *et al.,* 2006). Social mixing pattern which is common with alcohol users also contributes to a higher chance of being infected. Classen*et al.* (1999) demonstrated that 58%of the disease is being transmitted in a social gathering outside the households while drinking.

The habit of smoking tobacco and tuberculosis infection are amongst the greatest public health problems in the world and it has been projected that in 2020, the death related to tobacco would be approximately 8 million (Michael *et al.,* 2007). Smoking is one of the most common lifestyle habitamong men living in Mizoram. Tobacco smoking leads to disorder in respiratory tracts and cardiovascular system. Smokingof tobacco is critical for both cell mediated and humoral immune response; which is involved in cytokine release and inhibition (Phillips *et al.,* 2003). A study by Kolappan and Gopi (2002) showed the relation of smoking with TB with odds ratio (2.48) and the age adjusted odds ratio (2.24) which was considered statistically significant. A cross sectional study shows that current or ex-smokers had a higher

risk of infection than for those who do not smoke (Boon *et al.,* 2004). Various studies have been conducted, which uses different approaches to understand the relation between smoking and the disease (Michael *et al.,* 2007). It has also been thought that nicotine present in tobacco in some way lowers the immunity of a person, thus developing the disease.

In this study, gene variants have been identified which makes an individual susceptible to infection with tuberculosis. Through whole exome sequencing, gene variants which are known to be associated with the disease across different populations has also been identified in Mizo population. HGV&TB database reports 307 variants in 98 genes, out of which 101 are in the exonic region and only 30 variants are analysed to be pathogenic in association to tuberculosis. This study shows the presence of 16 such variants in Mizo population, which might make the population pre-disposed to *M.tuberculosis* infection. The second part of WES analysis also discovered variants that has not been reported in any population and are known to be pathogenic for the disease.

PACRG (rs10945890) has been found to be associated with TB in Russian population (Bragina *et al.,* 2016). In Mizo Population a variant (A/G) has been observed which causes an amino acid change from glutamate in 291 position to Glycine (E291G). This mutation introduces a more hydrophobic residue at this position that can result in loss of hydrogen bonds and/or disturb correct folding of the protein. Most of the proteins encoded by these genes are involved in immune signalling and are responsible for the effectiveness of immune reactions to the invasion of the pathogen. The variant of PACRG in the current study has not yet been found to be associated with TB.

TRIM is a tyrosine phosphatase PtpA- interacting host protein that restricts the survival of *Mtb* in macrophages. This gene acts as restriction factor blocking bacterial replication in infected cells (McGuire *et al.,* 2015). TRIM27, has been demonstrated to restrict the survival of the bacteria by promoting innate immune responses and cell apoptosis. Interestingly, Mtb PtpA could antagonize TRIM27-promoted JNK/p38 MAPK pathway activation and cell apoptosis through competitively binding to the RING domain of TRIM27. This study suggests a potential tuberculosis treatment via targeting of the TRIM27-PtpA interfaces (Wang *et al.,* 2016). The present study also describes the presence of a variant in TRIM46 gene, which leads to an amino acid change of glycine to cystine(G2836C). The wild-type residue is the most flexible of all residues whose flexibility might be necessary for the protein's function. Mutation of this glycine into another residue can force the local backbone into an incorrect conformation and will disturb the local structure. The change in structure of the protein might result in a truncated Ptp protein required for killing the pathogen thus losing its function.

The present study has also identified variant in SLC27A3 (N148K) and TCIRG (G510R) genes. SLC27A3 which is a member of solute carrier family which functions as an exchanger for proton and divalent cation, altering the phagolysosomal environment, thus reducing the replication capability of the bacterium (Courville *et al.,*2000), might also play an important in host defence against *Mtb.* SLC27A3 has not yet been studied for its association with TB.

TCIRG which is a T-Cell, Immune Regulator 1, ATPase, H+ Transporting, Lysosomal Subunit seems to be directly involved in T-cell activation.T- cells are important to destroy cells that are infected with bacteria and play important roles in

all arms of immunity. Mutation in this gene might be responsible for host susceptibility against *Mtb* infection.

The first and only report for discovery of the association of SNPs with tuberculosis using exome sequencing was done by Bhattacharya *et al*. (2019) for Bengali population. They discovered novel Single nucleotide polymorphisms (SNPs) in sialic acid binding immunoglobulin-like lectins 15 (SIGLEC15) gene and Human Leucocyte Antigen (HLA-DRA) gene. SIGLEC15 are known to be expressed on macrophages associated with tumor and are also involved in signal transduction pathway. Whereas, HLA-DRA gene plays a defensive role against Mtb by presenting peptides for recognition by T lymphocytes on antigen-presenting cells.

Many studies have reported the association of HLA-DRB1 and HLA-DQB1 in Chinese populations (Wu *et al.,* 2013), Brazilian Amazon (Lima *et al.,* 2016), Vietnam (Goldfeld *et al.,* 1998), Poland (Dubaniewicz *et al.,* 2003), Asian Indians (Brahmajothi *et al.,* 1991), Cambodian (Goldfeld *et al.,* 1998), Indonesian (Bothamley *et al.,* 1989) gene with tuberculosis. However, its association with HLA-DRA has not been reported for any populations.

This is the first study conducted for sequencing of the mitochondrial DNA for tuberculosis patients. Non- coding regions of the genome also showed the presence of variants that were found only in the patients, thus giving insights into its role in susceptibility to infection. We could also identify gene variants in the coding region which were deleterious as predicted by the function prediction tools. To know the impact of these novel variants identified, we need to validate it using functional studies which would provide a basis for treatment of the disease and control TB burden.

The mutation in MT-ND2 gene (T119A) and MT-ND6 gene (Y165C) was predicted to be deleterious/damaging in Polyphen-2 and Provean.These variants upon structural function prediction shows that it is located near a highly conserved position. The mutation in these residues will cause an empty space in the core of the protein and cause loss of hydrogen bonds in the core of the protein and as a result disturb correct folding of the protein.

Single deletions of the mitochondrial DNA (mtDNA) were the first pathogenic mutations to be reported (Holt *et al.,* 1988). MtDNA mutation leads to metabolic, neurological and muscular disorder, thusacting as a cause in human disease (Saxena *et al.,* 2006). Deletion in mtDNA have been suggested to take part in mutagenesis (Lakshmanan *et al.,* 2011).

MtDNA mutations commonly occurs in different forms of cancers and may be functional in cancer detections (Jessie *et al*., 2001). Sequencing the coding regions of mtDNAin thyroid cancer may contribute in the thyroid tumorigenesis especially in the early stage of cancer (Abu-Amero *et al.,* 2004). Sequencing of the complete mtDNA of colorectal cancer found homoplasmic mutations shows its involvement in abnormal metabolic and apoptotic processes in cancer. Mitochondrial mutation leads to defect in oxidative phosphorylation and electron transport due to increase in reactive oxygen. This also result in deletion in mitochondrial genome. Thus when there is an increased accumulation of deleted mtDNA, the mitochondrial cell continues to produce reactive oxygen by-products which damages the constituents of cell (Polyak *et al.,* 1998).

The other studied genes were not found to be directly associated with TB in the current study; nevertheless, their analysis in other populations may still be fruitful given their functional importance in TB pathogenesis.

Moreover, this study can serve as a baseline data to understand the current status of TB prevalence in Mizoram and the genetic pre-disposition of the population. However, the study has several limitations since it relied on small number of sample size. Large population studies are required to overcome the limitation of sample size and encompass all the variables including association with socio-demographic factors and the presence of the variants related to TB.

Summary

The work done in this present study and its findings can be summarized as follows:

1. Demographic and life style factors studies revealed Sex (Male), usage ofalcohol and tobacco, and consumption of smoked foods, as risk factors associated with the prevalence of TB infection.
2. Whole ExomeSequencing (WES) revealed the presence of reported pathogenic genes such as MC3R, P2RX7, LAMP1, MBL2, TIRAP, VDR, SFTPA1, SP110, WIPI11, MACRO, P2RX7, SLC11A1, IL1B, IL10RA, SFTPA1, and SFTPA2 in Mizo population, Mizoram.
3. WES studies also revealed the presence of about 10-30 single nucleotide variants, analysed to be pathogenic and not reported for any population under study.
4. Whole mtDNA sequencing revealed the presence of a total of 86 variants in the non-coding region and a total of 91 variants in the coding region of the mitochondrial genome.
5. Out of the 91 variants in the coding region of the mitochondrial genome, 27 were non-synonymous and 64 were synonymous.
6. Out of the 27 non-synonymous variants, MT-ND2 g. A>G4824 p.T119A, found in two patients (TB-25 & TB-37),and MT-ND6 g. T>C14180 p.Y165C, found in one patient (TB-25) was predicted deleteriousin Polyphen-2 and Provean.

   The findings from this study can serve as a means to spread better and more efficient awareness program on TB spread and infection, and also to aware them about the risk factors such as tobacco usage and alcohol drinking as it’s a common practice among adult men in Mizoram.

References

Abba K, SudarsanamTD, Grobler L and Volmink J. (2008) Nutritional supplements for people being treated for active tuberculosis. *Cochrane Database of Systematic Reviews*, no. 4, Article ID CD006086.

Abu-Amero KK, Alzahrani AS, Zou M and Shi Y. (2005) High frequency of somatic mitochondrial DNA mutations in human thyroid carcinomas and complex I respiratory defect in thyroid cancer cell lines. Oncogene. 24, 1455–1460

Akira S, Ueamatsu S and Takeuchi S. (2006) Pathogen Recognition and Innate Immunity. Cell. 124(4); 783-801.

Andrews S. (2010) FastQC: a quality control tool for high throughput sequence data. Available online at: http:// www.bioinformatics. babraham.ac. uk/projects/ fastqc Ann Intern Med. 1463 ; 4054.

Ao Y, Wang X, Cao Z and Cheng X. (2016) Vitamin D receptor geneFokI polymorphisms and tuberculosis susceptibility: a meta-analysis.Arch Med Sci. 11:256.-269

Ates O, Musellim B, Ongen G and Sarikaya AT. (2008) Interleukin-10 and Tumor Necrosis Factor-α Gene Polymorphisms in Tuberculosis. J ClinImmunol (2008) 28: 232.

Baghdadi JE, OrlovaM and Alter A. (2006)An autosomal dominant major gene confers predisposition to pulmonary tuberculosis in adults. J Exp Med. 203:1679–1684.Ng SB, Buckingham KT, Lee C, Bigham AW, Dent KM, Huff CD, Shannon PT, Jabs EW, Nickerson DA, Shendure J and Bamshad MJ. (2010)Exome sequencing identifies the cause of a mendelian disorder. *Nature Genetics*. volume42, pages 30–35.

Bellamy R, Ruwende C and Corrah T. (2004) Variations in the NRAMP1 gene and susceptibility to tuberculosis in West Africans. N Engl J Med. 338:640–644.

Bellamy R. Identifying genetic susceptibility factors for tuberculosis in Africans: a combined approach using a candidate gene study and a genome-wide screen (2000). Clin. Sci. (London) 98: 245–250.

Bhattacharyya C, Majumder PP, and Pandit B. (2019) Anexome wide association study of pulmonary tuberculosis patients and their asymptomatic household contacts. Infection, Genetics and Evolution. 71: 76-81.

Bragina, Yu E, Babushkina, Nadezhda P, Garaeva, Anna F, Rudko, Alexey A, Tsitrikov, Dmitry Y. (2016) Impact of the polymorphism of the PACRG and CD80 genes on the development of the different stages of Tuberculosis infection. Iranian Journal of Medical Sciences. 6: 34827

Cao H, Wu H and LuoR. (2015) De novo assembly of a haplotype-resolved human genome. Nat Biotechnol.33:617–622.

Cervino AC, LakissS, and Sow O. (2002) Fine mapping of a putative tuberculosis-susceptibility locus on chromosome 15q11-13 in African families. Hum Mol Genet. ;11:1599–1603

Chandler MR, Bilgili EP, and Nancy D. Merner ND. (2016). A Review of Whole-Exome Sequencing Efforts Toward Hereditary Breast Cancer Susceptibility Gene Discovery. HUMAN MUTATION. Vol. 37, No. 9, 835–846.

Courville P, Chaloupka R and Cellier MFM. (2006) Re- cent progress in structure-function analy- ses of Nramp proton-dependent metal-ion transporters. Biochem Cell Biol. 84: 960-78.

Davila S, Hibberd ML and Hari DR. (2008) Genetic association and expression studies indicate a role of toll-like receptor 8 in pulmonary tuberculosis. PLoS Genet. 4:e1000218

Devi KR, Ramalingam B, Brennan PJ, Naravanan PR and Raja A (2001) Specific and earlydetection of IgG, IgA and IgM antibodies to *Mycobacterium tuberculosis* 38kDa antigen in pulmonary tuberculosis. Tuberculosis. 81 (3): 249-253.

Dittrich N, Berrocal-Almanza LC, and Thada S. (2015) Toll-like receptor 1 variations influence susceptibility and immune response to Mycobacterium tuberculosis. Tuberculosis95: 328–335.

Dobler CC, Flack JR and Marks GB. (2012) Risk of tuberculosis among people with diabetes mellitus: an Australian nationwide cohort study. *BMJ Open*; 2:e000666.

Dodd P J, Looker C and Plumb I D (2016)Age- and sex-specific social contact patterns and incidence of Mycobacterium tuberculosis infection. Am J Epidemiol; 183: 156–166.

Drury S, Williams H, Trump N, Boustred C, Lench N, Scott RH and Chitty LS. (2015) Exome Sequencing for prenatal diagnosis of foetuses with sonographic abnormalities. Prenatal Diagnosis. 35:10.

Du J, Han J, Li X, Zhang Y, Li H and Yang S. (2015) IL-17 genepolymorphisms in the development of pulmonary tuberculo-sis.Int J ClinExpPathol. 8: 3225–3229.

Duan HF, Zhou XH andMa Y. (2003) A study on the association of 3'UTR polymorphisms of NRAMP1 gene with susceptibility to tuberculosis in Hans. ZhonghuaJieHeHeHuXiZaZhi. 26:286–289.

Ganz T and Nemeth E. (2015) Iron homeostasis in host defence and inflammation. Nat Rev Immunol; 15: 500–510.

Gao PS, Fujishima S and Mao XQ. (2000) Genetic variants of NRAMP1 and active tuberculosis in Japanese populations. International tuberculosis genetics team. Clin Genet.58:74–76

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

Greenwood CM, Fujiwara TM, BoothroydLJ, et al. Linkage of tuberculosis to chromosome 2q35 loci, including NRAMP1, in a large aboriginal Canadian family. Am J Hum Genet. 2000;67:405–416.

Habegger de Sorrentino A, Pardo R and Marinic K. (2014) KIR-HLA class i and pulmonary tuberculosis in the Amerindian population in Chaco, Argentina.EnfermInfeccMicrobiolClin. 32:565–569.

Harries AD, Lin Y, Satyanarayana S, Lönnroth K, Li L, Wilson N, Chauhan LS, Zachariah R, Baker MA, Jeon CY, Murray MB, Maher D, Bygbjerg IC, Enarson DA, Billo NE and Kapur A. (2011) The looming epidemic of diabetes-associated tuberculosis: learning lessons from HIV-associated tuberculosis. The International Journal of Tuberculosis and Lung Disease. 15 (11): 1436-1445.

Holt IJ, Harding AE, Morgan-Hughes JA. (1988) Deletions of muscle mitochondrial DNA in patients with mitochondrial myopathies. Nature. 331:717–19.

Horton KC,MacphersonP,HoubenRMGJ,White G and Corbett E L. (2016) Sex differences in tuberculosis burden and notifications in low- and middle-income countries: a systematic review and meta-analysis. PLOS Med; 13(9): e1002119.

Huang L, Liu C and Liao G. (2015) Vitamin D receptor gene foki polymorphism contributes to increasing the risk of tuberculosis: an update meta-analysis. Medicine (Baltimore). 94:e2256.

Imtiaz S, Shield KD, Roerecke M, Samokhvalov AV, Lönnroth K and Rehm A. (2017) Alcohol consumption as a risk factor for tuberculosis: meta-analyses and burden of disease. European Respiratory Journal. 50: 1700216.

India TB Report 2018. Revised National TB control programme. Annual status report. Central TB division. Ministry of Health and Family Welfare, New Delhi.

Jafari M, Nasiri MR, and Sanaei R.(2016) The NRAMP1, VDR, TNF-alpha, ICAM1, TLR2 and TLR4 gene polymorphisms in Iranian patients with pulmonary tuberculosis: A case-control study. Infect Genet Evol.39: 92–98.

Kardum LJ, Etokebe GE, Lederer P, Balen S, and Dembic Z. (2015) Genetic polymorphisms in the Toll-like receptor 10, interleukin (IL)17A and IL17F genes differently affect the risk for tuberculosis in Croatian population. Scand J Immunol. 82: 63–69.

Kim JH, Lee SY and Lee SH. (2003) NRAMP1 genetic polymorphisms as a risk factor of tuberculous pleurisy. Int J Tuberc Lung Dis. 7:370–375

Kinnear C, Hoal EG, Schurz H, Helden PDV and Moller M. (2017) The role of human host genetics in tuberculosis resistance. Expert Review of Respiratory Medicine. 11:9, 721-737.

Koboldt D, Zhang Q, Larson D, Shen D, McLellan M, Lin L, Miller C, Mardis E, Ding L, and Wilson R. (2012). VarScan 2: Somatic mutation and copy number alteration discovery in cancer by exome sequencing Genome Research DOI: 10.1101/gr.129684.111 URL: http://varscan.sourceforge.net

Kolappan C and Gopi P. (2002) Tobacco smoking and Pumonary tuberculosis. Thorax. 56(11): 964-966

Lakshmanan LN, Gruber J, Halliwell B, Gunawan R. (2012) Role of Direct Repeat and Stem-Loop Motifs in mtDNA Deletions: Cause or Coincidence?. PLoS ONE 7(4): e35271.

Landrum MJ, Lee JM, Riley GR, *et al.* ClinVar: public archive of relationships among sequence variation and human phenotype.(2014)Nucleic Acids Res;42(Database issue):D980–5.

Lee SW, Lin CY, Chuang TY, Huang HH, Kao YH, and Wu LS. (2016) SNP rs4331426 in 18q11.2 is associated with susceptibility to tuberculosis among female Han Taiwanese. J MicrobiolImmunol Infect. 49: 436–438.

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

Li H, Zhang T and Huang Q. (2006) Meta-analysis on NRAMP1 gene polymorphisms and tuberculosis susceptibility in East-Asia population. Zhonghua Liu Xing Bing XueZaZhiZhonghuaLiuxingbingxueZazhi.;27:428–432.

Li H. (2013) Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. arXiv:1303.3997v1.

Li X, Yang Y, Zhou F. (2011)SLC11A1 (NRAMP1) polymorphisms and tuberculosis susceptibility: updated systematic review and meta-analysis. PloS One. 6:e15831

Liu W, Cao WC and Zhang CY. (2004)VDR and NRAMP1 gene polymorphisms in susceptibility to pulmonary tuberculosis among the Chinese Han population: a case-control study. Int J Tuberc Lung Dis. 8:428–434.

Lönnroth K,Jaramillo E, Williams BG, Dye C and Raviglione M. (2008) Drivers of tuberculosis epidemics: The role of risk factors and social determinants. Social Science & Medicine. 68 (12): 2240-2246.

Lyadova I. (2010) Inflammation and Immunopathogenesis of Tuberculosis Progression. www.intechopen.com.

Mahasirimongkol S, YanaiH and Nishida N.(2009)Genome-wide SNP-based linkage analysis of tuberculosis in Thais. Genes Immun. 10:77–83.

McGuire VA, & Arthur JS. (2015). Subverting Toll-Like Receptor Signaling by Bacterial Pathogens. Front Immunol6, 607, 10.3389/fimmu.2015.00607.

Meilang Q, Zhang Y, Zhang. (2012) Polymorphisms in the SLC11A1 gene and tuberculosis risk: a meta-analysis update. Int J Tuberc Lung Dis.16:437-446.

Menzies D, Pai M and Comstock G (2008) Meta-analysis: new tests for the diagnosis of latent tuberculosis infection: areas of uncertainty and recommendations for research.

Meyer CG and Thye T. (2014) Host genetic studies in adult pulmonary tuberculosis. SeminImmunol.

Meyer CG, Reiling N and Ehmen C. (2016) TLR1 variant H305Lassociated with protection from pulmonary tuberculosis.PLoS ONE. 11: e0156046.

Möller M, Kinnear CJ, Orlova M, Kroon EE, Helden PD, Schurr E and Hoal EG. (2018). Genetic resistance to *Mycobacterium tuberculosis* infection and disease. Front Immunol. 27;9:2219.

Möller M, Wit E andHoalEG. (2010) Past, present and future directions in human genetic susceptibility to tuberculosis. FEMS Immunol Med Microbiol.;58:3–26

Motulsky AG. (1960) Metabolic polymorphisms and the role of infectious diseases in human evolution. Hum Biol. 32:28–62

North RJ and Jung YJ. (2004) Immunity to tuberculosis. Annual Review of Immunology. Vol. 22:599-623.

Pacheco AG, Cardoso CC and MoraesMO. (2008)IFNG +874T/A, IL10-1082G/A and TNF −308G/A polymorphisms in association with tuberculosis susceptibility: a meta-analysis study. Hum Genet.;123:477–484.

Palodhi A, Ghosh S, Biswas NK, Basu A, Majumder PP, Maitra A and India Project Team of the International Cancer Genome Consortium. (2019) Profiling of Genomic Alterations of Mitochondrial DNA in Gingivobuccal Oral Squamous Cell Carcinoma: Implications for Disease Progress.

Polyak K, Li Y, Zhu H, Lengauer C, Willson JKV, Markowitz SD, Trush MA, Kinzler KW and Vogelstein B. (1998). Somatic mutations of the mitochondrial genome in human colorectal tumours. Nature Genetics Vol 20.

Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces.BMC Bioinformatics. 2010 Nov 8;11(1):548. DOI: 10.1186/1471-2105-11-548. PubMed: 21059217.

Ravikumar M, Dheenadhayalan V, Rajaram K, Lakshmi SS, Kumaran PP, Paramasivan CN and Balakrishnan K. (1999) Associations of HLA-DRB1, DQB1 and DPB1 alleles with pulmonary tuberculosis in south India. Tuber Lung Dis 79: 309–317.

Robinson JT, Thorvaldsdóttir H, Winckler W, Guttman M, Lander ES, Getz G, Mesirov JP. (2011) Integrative genomics viewer. Nat Biotechnol. 29(1):24-6.

Ryu S, Park YK and Bai GH. (2000) 3'UTR polymorphisms in the NRAMP1 gene are associated with susceptibility to tuberculosis in Koreans. Int J Tuberc Lung Dis. 4:577–580.

Sahajpal R, Kandoi G,Dhiman H, Raj S, Open Source Drug Discovery Consortium, Scaria V, Bhartiya D andHasija Y. (2014) HGV&TB: a comprehensive online resource on human genes and genetic variants associated with tuberculosis.

Salie M, Daya M, and Lucas LA. (2015. Association of toll-like receptors with susceptibility to tuberculosis suggests sex-specific effects of TLR8 polymorphisms. Infect Genet Evol34: 221–229.

Selvaraj P, Narayanan PR and Reetha AM. (1999) Association of functional mutant homozygotes of the mannose-binding protein gene with susceptibility to pulmonary tuberculosis in India. Tubercle and Lung disease. 79 (4): 221-227.

Shaw MA, Collins A and Peacock CS. (1997) Evidence that genetic susceptibility to Mycobacterium tuberculosis in a Brazilian population is under oligogenic control: linkage study of the candidate genes NRAMP1 and TNFA. TuberLung Dis. 78:35–45

Shin HD, Park BL, Kim LH, Cheong HS, Lee IH and Park SK. (2005) Common interleukin 10 polymorphism associated with decreased risk of tuberculosis. Experimental & Molecular Medicine 37, 128–132.

Shin JG, Park BL and Kim LH. (2015) Association study of poly-morphisms in interferon-gamma receptor genes with the riskof pulmonary tuberculosis.Mol Med Rep:12:1568–1578

Slama K, Chiang CY, Enarson DA, Hassmiller K, Fanning A, Gupta P and Ray C. (2007) Tobacco and tuberculosis: a qualitative systematic review and meta-analysis The International Journal of Tuberculosis and Lung Disease. 11 (10):2001049-1061(13).

Stead WW, SennerJW andReddick WT. (1990) Racial differences in susceptibility to infection by Mycobacterium tuberculosis. N Engl J Med. 322:422–427

Stein CM, ZalwangoS, Malone LL, et al. Genome scan of M. tuberculosis infection and disease in Ugandans. PLoS One. 2008;3:e4094

Sui G, Zhou S, Wang J, Canto M, Lee EE, Eshleman JR, Montgomery EA, Sidransky D, Califano JA and Maitra A. (2006) Mitochondrial DNA mutations in preneoplastic lesions of the gastrointestinal tract: A biomarker for the early detection of cancer. Molecular Cancer. 5:73-82.

Tong HV, Velavan TP, Thye T, and Christian G. Meyer CG. (2017). Human genetic factors in TB. Tropical Medicine and International Health volume 22 no 9 pp 1063–1071.

Velez DR, Hulme WF and Myers JL. (2009)NOS2A, TLR4, and IFNGR1 interactions influence pulmonary tuberculosis susceptibility in African-Americans. Hum Genet. 126:643–653

Wang J,Teng JLL, Zhao D, Ge P, Li B, Woo PCY, and Liu CH. (2016) The ubiquitin ligase TRIM27 functions as a host restriction factor antagonized by Mycobacterium tuberculosis PtpA during Mycobacterial Infection. Scientific Reports. 6:34827.

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

WHO. Global tuberculosis report 2018 [Internet]. WHO. Available from: http://www.who.int/tb/publications/global_report/en/

Wu L, Hu Y, Li D, Jiang W, and Xu B. (2015) Screening toll-like receptor markers to predict latent tuberculosis infection and subsequent tuberculosis disease in a Chinese population. BMC Med Genet2015: 16: 19.

Yang Y, Muzny DM, Reid JG, Bainbridge MN, Willis A, Ward PA, Braxton A, Beuten J, Xia F, Niu Z, Hardison M and Pearson R. (2013) Clinical Whole-Exome sequencing for the Diagnosis of Mendelian Disorders. N Engl J Med 369:1502-1511.

Yi YX, Han JB, Zhao L, Fang Y, Zhang YF and Zhou GY. (2015) Tumor necrosis factor alpha gene polymorphism contributesto pulmonary tuberculosis susceptibility: evidence from ameta-analysis.Int J ClinExp Med. 8: 20690–20700.

Zhao Y, Bu H,andHong K. (2015) Genetic polymorphisms of CCL1 rs2072069 G/A and TLR2 rs3804099 T/C in pulmonary or meningeal tuberculosis patients. Int J ClinExpPathol8: 12608–12620.

MARY VANLALHRUAII TONSING
M.Phil. Research Scholar
Department of Biotechnology
Mizoram University
Aizawl -796 004, Mizoram, India

E-mail: marytonsing12@gmail.com
Mobile: +91 - 9774656098

I have attained my Bachelor's and Master's degree in Biotechnology. I am pursuing my M.Phil. degree in the Department of Biotechnology, Mizoram University under the supervision of Prof. N. Senthil Kumar.

Educational Qualification

Degree	**Board**	**Division**	**Month and Year of Joining**	**Month and Year of Completing**
HSLC (Class 10)	Meghalaya Board of School Education	I	February 2008	March 2009
HSSLC (Class 12)	Indian School Certificate	I	June 2009	March 2011
B.Sc. Biotechnology	North Eastern Hill University	I	June 2011	March 2014
M.Sc. Biotechnology	Mizoram University	I	August 2014	July 2016
M.Phil	Mizoram University	Dist.	August 2017	

Awards

Awarded 7th Rank in B.Sc. (Biotechnology), North Eastern Hill University Examination, 2014.

Skills developed

DNA/RNA Isolation,Gel electrophoresis, PCR, Sanger Sequencing and Next Generation Sequencing, Real Time-PCR, Bioinformatics tools and methods.

Conference attended

Poster presentation on **"Polymorphism in aldehyde dehydrogenase-2 gene associated with risk of tuberculosis in Mizo population",** in International Conference on Biodiversity, Environment and Human Health : Innovations and Emerging Trends (BEHIET 2018), Organized by School of Life Sciences, Mizoram University, Aizawl, India and Association of Biotechnology and Pharmacy(ABAP), India during 12 – 14 November 2018.

Hands-on Trainings attended

1. 3rd NER Hands-on Training Workshop on "Gene Cloning, Protein Biochemistry, Structure Biology and Bioinformatics" organised by, Advanced Centre for Treatment, Research & Education in Cancer, Khargarh, Navi Mumbai, India, during 04 – 15 February 2019.
2. 3rd Advanced NER Research Training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics-Driven Approaches" held at National Institute of Biomedical Genomics, Kalyani from 23 – 31 July 2018.
3. 3rd NER Research Training Workshop on "Understanding Human Disease and Improving Human Health Using Genomics-Driven Approaches" held at Department of Biotechnology, Mizoram University, organised by National Institute of Biomedical Genomics, Kalyani from 19 – 24 November 2017

Workshops attended

1. Workshop on “Introduction to basic and advanced biomedical approaches for enhancing QOL in aging societies” held at Biomedical Research Institute, AIST, Tsukuba Science City, Japan, from 14 – 21 October, 2018
2. Workshop on “The Concept and Application of Genomics in Clinical Medicine” in Continuing Medical Education Course held on 11 August 2018 at Civil Hospital Aizawl, conducted by, CSIR- Institute of Genomics and Integrative Biology (CSIR-IGIB), New Delhi.
3. Workshop on “Statistical and Computing Methods for Life-Science Data Analysis” held during 05 – 10 March, 2018 Jointly organized by Biological Anthropology Unit, Indian Statistical Institute, Kolkata and Department of Botany, Mizoram University, Aizawl.
4. National Workshop on “Sequencing and Fragment Analysis” held during 26 Feb – 01 March, 2018 organized by Advanced Level State Biotech Hub Facility, Department of Biotechnology, Mizoram University.
5. National Workshop on :Statistical Methods in Biological research” held during 03 – 05 November, 2017 organized by Bioinformatics Infrastructure Facility(BIF), Department of Biotechnology, Mizoram University.
6. International Workshop on “Molecular Phylogeny and Next-Generation Sequencing” held at Department of Biotechnology, Jointly organized by Global Initiative for Academic Networks and Mizoram University, during 19 – 28 June 2017.
7. International Workshop on “Molecular Entomology” held at Department of Zoology, Jointly organized by Global Initiative for Academic Networks and Mizoram University, during 19 – 28 June 2017.

Paper publications

Papers under construction

Mary VanlalhruaiiTonsing, Christine VanlalbiakdikiSailo, Lily Chhakchhuak, Zothansanga, NachimuthuSenthil Kumar. Characterization of mtDNA variation of Tuberculosis patients in Mizo population. Mitochondrial DNA. ISSN: 1940-1736. Impact factor (1.760).

Mary VanlalhruaiiTonsing, Christine VanlalbiakdikiSailo, Lily Chhakchhuak, Zothansanga, NachimuthuSenthil Kumar. An exome sequencing approach to understand the underlying genetic variants associated with susceptibility or resistance to tuberculosis in Mizo population. Infection, Genetics and Evolution.

Papers published

Mary V Tonsing, SouvikGhatak, Freda Lalrohlui, NachimuthuSenthil Kumar, John Zohmingthanga (2018). Five years Record on Cancer Incidence from a Diagnostic Centre in Mizoram, Northeast India. *Cancer Health Disparities.*ISSN: 2573 – 9530.

DhaneshwareeAsem, Vincent Vineeth Leo, Ajit Kumar Passari,**Mary VanlalhruaiiTonsing**, J. Beslin Joshi ,SivakumarUthandi, AbeerHashem, ElsayedFathiAbd Allah, Bhim Pratap Singh (2017). Evaluation of gastrointestinal bacterial population for the production of holocellulose enzymes for biomass deconstruction*Plosone*12(10):e0186355.doi:10.1371/journal.pone.0186355. **(*Impact factor*: 2.8).**