HSI Health Sciences Investigations Journal

share

Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

ISSN Online 2704-4890 | ISSN Print 2720-7609

Online first publication

Original Research Article

HSI Journal (2026) Volume 8 (Issue 2):1476-1489. https://doi.org/10.46829/hsijournal.2026.4.8.2.1476-1489

Open Access

Pathogenic variants in *Mycobacterium tuberculosis* susceptibility genes: a cross-continental bioinformatics analysis

Lalu M IRHAM [1], Wirawan ADIKUSUMA [2], Danang P AMUKTI [3*], Sabiah KHAIRI [2], Petrina T PHILOTHRA [4], Ichtiarini N SANTRI [5], Isom HILMI [6], Rockie CHONG [7], Daraporn RUNGPRAI [8], Satriya PRANATA [9], Baik H RISPAWATI [10], MURIYATI [11], Ilker ATES [12]

[1] Faculty of Pharmacy, Universitas Ahmad Dahlan, Yogyakarta, Indonesia; [2] Research Center for Computing, Research Organization for Electronics and Informatics, National Research and Innovation Agency (BRIN), Cibinong Science Center, Cibinong, Indonesia; [3] Department of Pharmacy, Faculty of Health Sciences, Alma Ata University, Yogyakarta, Indonesia; [4] Department of Rehabilitation Medicine, General Hospital Yogyakarta City, Yogyakarta, Indonesia; [5] Faculty of Public Health, Universitas Ahmad Dahlan, Yogyakarta, Indonesia; [6] Department of Physics, Faculty of Applied Science and Technology, Universitas Ahmad Dahlan, Yogyakarta, Indonesia; [7] Department of Chemistry and Biochemistry, University of California, Los Angeles, USA; [8] Faculty of Pharmacy, Silpakorn University, Nakhon Pathom, Thailand; [9] Department of Nursing, Faculty of Nursing and Health Sciences, Universitas Muhammadiyah Semarang, Semarang, Central Java, Indonesia; [10] Institut Kesehatan YARSI Mataram, Mataram, Indonesia; [11] Stikes Panrita Husada Bulukumba, South Sulawesi, Indonesia; [12] Faculty of Pharmacy, Department of Toxicology, Ankara University, Yenimahalle, Ankara, Turkey

Received December 2025; Revised January 2026; Accepted March 2026

Abstract

Background: This study investigated genetic factors influencing susceptibility to *Mycobacterium tuberculosis* (M. tuberculosis) infection across diverse populations to inform precision TB interventions.

Objective: This study investigated genetic factors influencing susceptibility to *M. tuberculosis* infection across diverse populations to inform precision TB interventions.

Methods: A retrospective review of LAF records, training logs, publications, and collaborations (2000–2025) was conducted. Data were summarised using descriptive statistics, with research models and training outputs organised thematically.

Results: The LAF produced over 20,000 Specific Pathogen Free (SPF) rodents, supporting >75 ethically approved projects in communicable and non-communicable disease research. It trained >170 researchers in animal science and contributed to >500 peer-reviewed publications. Key disease models developed included Buruli ulcer, malaria, diabetes, epilepsy, benign prostate hyperplasia (BPH), and wound healing. The facility's output is regionally significant, with 60% of supported projects involving international collaborations. Continuous upgrades have enhanced biosafety and welfare standards, as well as ISO/IEC 17025:2017-aligned operations.

Conclusion: These variants highlight population-specific genetic risks for TB and potential for personalised prevention strategies. Further research into host-pathogen interactions is needed to optimise TB control.

Keywords: Bioinformatics, tuberculosis, drug repurposing, genomic variants, susceptibility genes, therapeutic targets

Cite the publication as Irham LM, Adikusuma W, Amukti DP and Khairi S et al. Pathogenic variants in TB susceptibility genes: A cross-continental bioinformatics analysis. HSI Journal 8 (2): 1476-1489. https://doi.org/10.46829/hsijournal.2026.4.8.2.1476-1489

INTRODUCTION

Tuberculosis (TB) remains a significant global health challenge, currently recognised as the leading infectious disease killer, surpassing COVID-19. According to the latest reports from the World Health Organisation (WHO), TB caused approximately 1.25 million deaths in 2023, with an estimated 10.8 million new cases reported worldwide [1]. Statistics from the Global TB Report 2021 regarding the incidence and impact of TB indicated that out of 824,000 estimated cases, only 393,323 were reported to health authorities. This discrepancy between estimated and notified cases highlights challenges in diagnosis and reporting. The report noted 3,110 deaths due to TB, reflecting the significant lethality of the disease

* *Corresponding author*
Email: amuktidanang@gmail.com

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

and its status as a major public health concern. These statistics illustrate the ongoing burden of TB worldwide, highlighting both progress in treatment outcomes and persistent challenges in achieving comprehensive coverage and timely diagnosis. The data underscore the need for continued efforts to improve TB detection, treatment access, and overall public health strategies to combat this infectious disease effectively [2].

TB infection is a complex disease influenced by a variety of multifactorial factors, including advanced age, family history, ancestry, and other genetic and environmental determinants. Understanding the aetiology of TB requires an exploration of both exogenous and endogenous risk factors that contribute to susceptibility and disease progression [3]. The incidence of tuberculosis varies significantly across ethnic groups and geographic regions due to a complex interplay of genetic predispositions, cultural practices, environmental factors, and socioeconomic conditions. Understanding these variations is vital to developing targeted public health strategies to reduce TB transmission and improve treatment outcomes across diverse populations. Continued research into the interactions between host genetics and pathogen characteristics will further elucidate the mechanisms underlying these disparities in TB incidence globally [4,5]. Genome-wide association studies (GWAS) have emerged as a powerful tool for identifying genetic variants associated with tuberculosis (TB) susceptibility across diverse populations [6]. The application of GWAS to tuberculosis research has identified significant genetic variants associated with disease susceptibility across diverse ethnic groups and geographic regions. These studies emphasise the importance of considering population diversity when investigating the genetic basis of TB, as well as the need for tailored public health strategies that address these disparities. Continued research using GWAS methodologies will likely uncover additional loci and enhance our understanding of the complex interactions between genetics and environmental factors that influence TB risk worldwide [6,7].

Despite the advancements in GWAS methodologies, the identification of pathogenic variants for TB remains limited. Most known single-nucleotide polymorphisms associated with TB susceptibility are already well documented, and while new variants are being discovered, their functional implications and contributions to disease pathogenesis require further exploration [7,8]. GWAS has proven to be a valuable approach for uncovering genetic variants associated with tuberculosis susceptibility and resistance; however, the number of identified pathogenic variants remains limited. Continued research is essential to elucidate the functional roles of single-nucleotide polymorphisms (SNPs), particularly in diverse populations worldwide affected by TB. Understanding these genetic factors will enhance our ability to develop targeted interventions and improve outcomes for individuals at risk for this infectious disease [8]. Despite the advancements in GWAS methodologies, the identification of pathogenic variants for TB remains limited. Functionally, genes such as HLA-DQA1 are critical to antigen presentation pathways that initiate adaptive immune responses against M. tuberculosis infection. Similarly, LGSN, HLA-DRB5, and RGS12 are involved in processes relevant to immune regulation and cellular signalling, which may influence host susceptibility and response to TB infection [9,10].

However, current studies have generally focused on single populations or limited geographic regions, underscoring a critical gap in cross-continental comparisons of pathogenic variants across diverse populations. Our study advances prior GWAS findings by applying a comprehensive bioinformatics approach to integrate and analyse epidemiological variant data from multiple continents, thereby elucidating population-specific and shared genetic factors that contribute to global TB susceptibility. The utilisation of bioinformatics in genomic databases offers substantial opportunities for identifying pathogenic variants associated with tuberculosis. By leveraging advanced computational tools and large-scale genomic data, researchers can enhance our understanding of TB's genetic underpinnings, improve diagnostic accuracy, and inform treatment strategies tailored to individual patient needs. As biomedical science continues to evolve, these bioinformatics approaches will be critical in addressing the global challenge posed by tuberculosis [11,12]. The integration of genomic data with transcriptomic and proteomic information through bioinformatics platforms enables a more comprehensive understanding of TB pathogenesis. This multi-omics approach can help identify not only genetic variants but also further demonstrate variant effects on gene expression and protein function, providing insights into disease mechanisms [12]. By identifying specific pathogenic variants associated with TB susceptibility or drug resistance, bioinformatics can play a pivotal role in developing personalised treatment strategies.

Understanding an individual's genetic details may allow healthcare providers to effectively practice personalised medicine. Therefore, our study aims to investigate epidemiological variant data across multiple continents for tuberculosis, utilising bioinformatics-based approaches.

MATERIALS AND METHODS

Prioritisation of genomic variants linked to tuberculosis susceptibility

Prioritising genomic variants related to TB susceptibility is an essential step in uncovering the genetic foundations of the disease. By focusing on significant variants identified through extensive research, scientists can deepen their understanding of TB pathogenesis and develop targeted strategies to address this global health issue. In our current study, we utilised GWAS database to identify variants associated with susceptibility to TB. We accessed GWAS catalog database on March 14, 2024. GWAS catalog serves as a publicly accessible resource that allows researchers to

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

investigate the connections between genetic variants and traits across various populations. The main objective of GWAS is to improve our comprehension of disease biology, providing a vast array of variants linked to phenotypic susceptibility [18]. The GWAS Catalog is a comprehensive resource for variant data, offering several significant advantages for researchers in genetics and genomics. The GWAS Catalog comprises a vast collection of variant-trait associations, with over 45,000 published GWAS studies covering more than 5,000 human traits [19]. This extensive dataset provides researchers with detailed information on genetic variants associated with various diseases and traits. The GWAS Catalog is frequently updated to include new studies and findings, ensuring users have access to the latest data in genetics. This continuous improvement helps maintain the relevance and utility of the database [20].

In our follow-up analysis, we prioritised genetic variants linked to the genes of interest using the HaploReg (https://pubs.broadinstitute.org/mammals/haploreg/haploreg.php) and Ensembl (https://www.ensembl.org/index.html?redirect=no) databases. HaploReg is an important resource that enables researchers to investigate annotations of non-coding genomic regions and to evaluate the functional effects of single-nucleotide polymorphisms (SNPs) within haplotype blocks. It provides valuable information on chromatin states, regulatory motifs, and evolutionary conservation across species, all of which are crucial for understanding how these variants may affect gene expression and susceptibility to disease [21]. Ensembl, conversely, provides a comprehensive genome browser that consolidates genomic data from various species, offering detailed annotations for genes, variants, and their potential functional effects. By utilising both HaploReg and Ensembl, we aimed to systematically assess and prioritise variants based on their anticipated biological significance and their possible involvement in tuberculosis susceptibility. This method enhances our capacity to identify variants that could clarify the genetic basis of TB and guide future research in this field [22,23].

The Ensembl database offers numerous advantages that make it a valuable resource for researchers in genomics and bioinformatics. Ensembl provides detailed and stable automatic annotations of various genomes, including gene predictions, regulatory regions, and functional elements [20]. This extensive annotation is fundamental for understanding the biological significance of the genomic sequence. Ensembl offers an interactive web interface that allows users to easily navigate and visualise genomic data. The customisable genome browser enables researchers to focus on specific genes or regions of interest, facilitating in-depth analysis [22]. HaploReg is a powerful tool designed for exploring annotations of non-coding genomic regions, particularly in relation to genetic variants and their potential functional impacts. HaploReg provides detailed annotations for non-coding variants, enabling researchers to investigate their potential regulatory roles in gene expression and disease susceptibility. This is particularly important given that many disease-associated variants are located in non-coding regions. The tool is particularly useful for post-GWAS analyses, helping researchers identify potentially causal variants within haplotype blocks associated with diseases and aids in prioritising variants for further functional studies [24].

Finally, we identified the genes with higher expression using the GTEx Portal (https://www.gtexportal.org/home/) across various tissues, including whole blood, lung, and liver (Figure 1). The GTEx portal facilitates the exploration of expression quantitative trait loci (eQTLs), which are genetic variants that influence gene expression levels. By prioritising genes with higher expression, researchers can better understand how specific SNPs may contribute to TB susceptibility or severity, particularly if those SNPs reside within regulatory regions that affect gene expression [18]. Identifying genes with higher expression in the GTEx portal across various tissues raises several important issues and opportunities that can significantly impact research and clinical applications related to tuberculosis (TB) and other diseases. Identifying genes with elevated expression levels in specific tissues can provide insights into their functional roles in disease processes. For TB, understanding which genes are highly expressed in the lungs or immune cells can shed light on the biological mechanisms underlying susceptibility and resistance to the infection.

Prioritising tuberculosis genes via GWAS and HaploReg

We assessed the variants that met the inclusion criteria for this study using a statistical significance threshold of a p-value < 5 x 10^-8, as outlined in the GWAS approach (https://www.ebi.ac.uk/gwas). We ensured that any duplicate SNPs were eliminated, allowing us to concentrate on the unique SNPs. After identifying the variants associated with TB, we then focused on variants encoded in the genes using HaploReg version 4.

Pathogenic SNP analysis via SNPnexus and HaploReg

After sorting and eliminating duplicates using Microsoft Excel, we identified 4 SNPs. These SNPs were then evaluated using SNPnexus (https://www.snp-nexus.org) by submitting batch queries for the 4 SNPs, selecting relevant annotation categories, including the HaploReg Database and 1000 Genome Population Data via Ensembl. The advantages of utilising HaploReg and Ensembl include a user-friendly interface and a comprehensive database of annotation fields that support batch searches and data visualisation, all without requiring extensive programming skills or significant computing resources from users. This makes these tools suitable for analysing and interpreting sequence variants across a wide array of biological applications.

The HaploReg Database facilitated the identification of SNP variants that influence protein changes associated with the disease, as demonstrated by missense mutations. Consequently, identifying these missense mutations

enables the assessment of the impact of single amino acid substitutions on protein function and structure. The data were collected on March 22, 2024, and after analysis, this study identified 4 SNPs predicted to be pathogenic variants.

Population distribution of SNPs via HaploReg and 1000 Genomes

The HaploReg v4 database was used to retrieve population-specific allele frequencies for the identified SNPs from the 1000 Genomes Project Phase 3 dataset. This database categorises global populations into five standardised superpopulations: African (AFR), Admixed American (AMR, encompassing Latin American populations of mixed Indigenous/Native American, European, and African ancestry), European (EUR), East Asian (EAS), and South Asian (SAS). This classification accurately represents South American genetic diversity within the AMR superpopulation while maintaining genetically meaningful continental comparisons. The database provided comprehensive reference data for human genetic variation, enabling precise estimation of continent-specific allele frequencies for our TB susceptibility variants.

Gene expression analysis across tissues using GTEx

The four genes corresponding to the four SNPs were analysed to examine gene expression across different tissues using the GTEx Portal (http://www.gtexportal.org/home/), with data extracted on March 25, 2024. The GTEx project serves as a large-scale resource that helps elucidate the complex patterns of genetic variation and gene regulation present in various human tissues. This includes analysing and interpreting the GWAS database for translational research. Consequently, using the GTEx Portal facilitated a thorough interpretation of this data across multiple tissues that may be relevant to various diseases.

This study evaluated bulk tissue gene expression for each gene, focusing on the top ten expressions sorted by log scale and median. The overall methodology was also applied in several bioinformatics studies related to chickenpox [25], diabetes mellitus [26], and Stevens-Johnson syndrome [27]. While the GTEx portal provides comprehensive bulk tissue expression data that aids in prioritising genes relevant to tuberculosis susceptibility, it is important to recognise that bulk tissue measurements represent averaged expression across heterogeneous cell populations. This may mask cell type-specific expression differences critical for precise mechanistic insights. Future integration of single-cell transcriptomic datasets would complement our findings by resolving the cellular context of variant effects.

RESULTS

Identification of TB-Associated Genomic Variants

Using the GWAS catalog database, we identified 5 unique pathogenic variants associated with TB susceptibility across various continents. After we removed duplicate variants, we identified 4 SNPs (Table 1). Next, we utilised HaploReg version 4 to match the variants encoded in the genes. Based on this analysis, we identified four variants

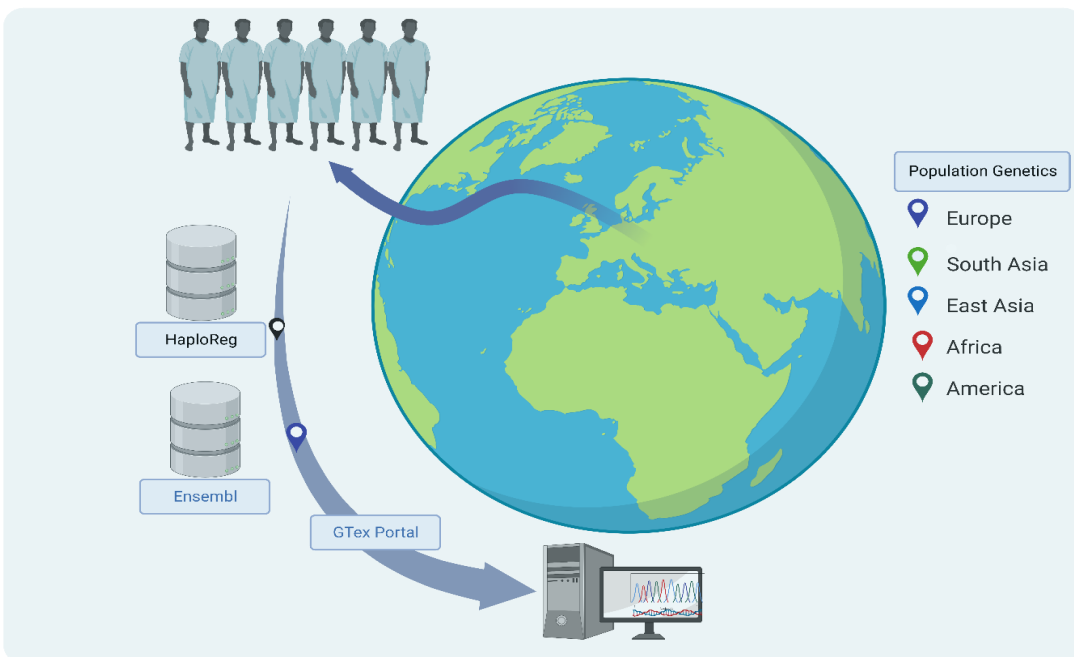

Figure 1. Schematic of custom bioinformatics pipeline for identifying tuberculosis disease susceptibility genes across multiple continents.

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

encoded by the genes. Our study identified four unique variants with missense mutations associated with *Mycobacterium tuberculosis* infection across multiple continents: rs35296992, which encodes LGSN; rs9272785, which encodes HLA-DQA1; rs41553512, which encodes HLA-DRB5; and rs2269497, which encodes RGS12. These SNPs were prioritised not only by a stringent p-value threshold but also because they are missense variants likely to impact protein function, observed across multiple continents and indicating broader population relevance, and involve genes known to play critical roles in immune responses against TB. Furthermore, allele frequency analysis confirmed their presence in diverse populations, strengthening the prioritisation of these pathogenic variants for TB susceptibility. This study highlighted that identifying pathogenic variants is crucial for identifying the most susceptible variants to TB.

Identification of TB-Associated Genes with missense mutations

According to HaploReg, we mapped the variants to the corresponding genes, with missense/nonsense mutations as non-synonymous changes resulting from a single base substitution to a different amino acid in the protein. In this step, we identified four variant genes with a missense mutation. LGSN amino acid changes p.Phe270Leu, HLA-DQA1 amino acid changes p.Ala210 Thr, HLA-DRB5 amino acid changes p.Val232Ile, RGS12 amino acid changes p.Asn1124Ser. The study involved mapping identified variants onto their corresponding genes, specifically identifying missense and nonsense mutations. Missense mutations are a type of non-synonymous change where a single base substitution results in the coding of a different amino acid in the protein, potentially altering its function. In this analysis, four specific variant genes were identified that contain missense mutations. These variants are. 1) The variant gene LGSN leads to a change from phenylalanine (p.Phe270) to leucine (Leu). 2) The variant gene HLA-DQA1 results in an alanine (p.Ala210) being replaced by threonine (Thr). 3) The variant gene HLA-DRB5 the mutation that changes valine (p.Val232) to isoleucine (Ile). 4) The variant gene RGS12 involves a substitution of asparagine (p.Asn1124) with serine (Ser).

Missense mutations can significantly impact protein function depending on the nature of the amino acid change. For instance, if the substituted amino acid has different properties (such as charge or hydrophobicity), it may affect protein folding, stability, or interactions with other molecules. Identifying these mutations is critical because they may influence susceptibility to M. tuberculosis infection by affecting immune response mechanisms. Understanding how these genetic changes alter protein function can help elucidate the biological pathways involved in TB susceptibility [28,29]. The aforementioned finding implies that this mapping and identification process is part of a broader research effort to understand the genetic factors underlying TB susceptibility across populations. This is crucial for developing targeted interventions and therapies [30,31].

Distribution of pathogenic TB-Associated SNP variants across populations

This finding emphasises the SNPs: rs35296992, which encodes LGSN; rs9272785, which encodes HLA-DQA1; rs41553512, which encodes HLA-DRB5; and rs2269497, which encodes RGS12. For rs35296992 (LGSN), it was observed that Europeans had the highest distribution among the three SNPs associated with TB susceptibility, followed by populations from America, Southeast Asia, and East Asia. In the case of rs9272785 (HLA-DQA1), the highest distribution was observed in Africans, followed by Southeast Asians, Europeans, East Asians, and Americans. For rs41553512 (HLA-DRB5), the highest frequency was observed in the Americas, followed by East Asians, Europeans, Africans, and Southeast Asians. Lastly, for rs2269497 (RGS12), the highest distribution was also observed in Americans, followed by East Asians, Europeans, Southeast Asians, and Africans. According to Table 2, Americans exhibited a higher distribution of SNPs for two variants: rs41553512 and rs2269497. In contrast, the highest distribution for Africans was observed with rs9272785, while Europeans showed the highest distribution for rs35296992.

Tissue-specific expression of TB-Associated variants

Following this, we investigated whether the TB-associated genes we identified included cis-expression quantitative trait loci (eQTL) in whole blood, liver and lung tissue. To achieve this, we integrated this information with the knowledge that functional variants regulate protein expression. Specifically, we determined that TB-associated genes exhibit enhanced expression in blood and lung tissue, as *Mycobacterium tuberculosis* predominantly affects these organs. To determine gene expression levels in human tissues linked to TB susceptibility, we utilised the Genotype-Tissue Expression (GTEx) portal database at [http://www.gtexportal.org/home](http://www.gtexportal.org/home/. The GTEx database is renowned for its comprehensive documentation of gene-tissue correlations and molecular phenotypes across multiple reference tissues [32]. Our analysis indicated that the four genes rs35296992 encoding LGSN (Figure 2A), rs9272785 encoding HLA-DQA1 (Figure 2B), rs41553512 encoding HLA-DRB5 (Figure 2C), and rs2269497 encoding RGS12 (Figure 2D), are expressed differently in lung, liver, and whole blood tissues.

The LGSN gene was expressed at higher levels in the liver than in the lung or whole blood. HLA-DQA1 gene has the highest expression seen in lungs relative to other tissues. HLA-DRB5 shows higher expression in lungs, followed by whole blood and liver. RGS12 gene also showing higher expression in lungs, followed by liver and whole blood. These findings underscore the importance of identifying pathogenic variants that drive differential gene expression across tissues relevant to TB pathology.

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

Correlation of allele frequency of variants with the incidence of TB disease in the world

Based on the tuberculosis (TB) incidence graph (Figure 3), it is evident that there are significant differences in the number of new TB cases across continents. Europe shows the lowest number of new cases, possibly due to the success of public health programs and good access to health services. On the other hand, Africa has a very high incidence of TB, which may be due to factors such as poverty, limited access to health services, and high HIV prevalence. The Americas show a lower number of new TB cases than Africa but higher than Europe, reflecting the challenges in controlling TB in the region. Meanwhile, Asia has the highest incidence of TB, possibly due to its very large population, rapid urbanisation, and challenges in the health care system in some Asian countries. These data are critical to understanding the different burdens of TB around the world and highlight the need for specific TB control strategies for each region. To strengthen the correlation analysis, we included Pearson's correlation coefficients (r) and p-values in Figure 4. The results show moderate-to-strong positive correlations between allele frequencies and TB incidence for rs35296992 (r = 0.80, p = 0.203), rs9272785 (r = 0.93, p = 0.075), and rs2269497 (r = 0.79, p

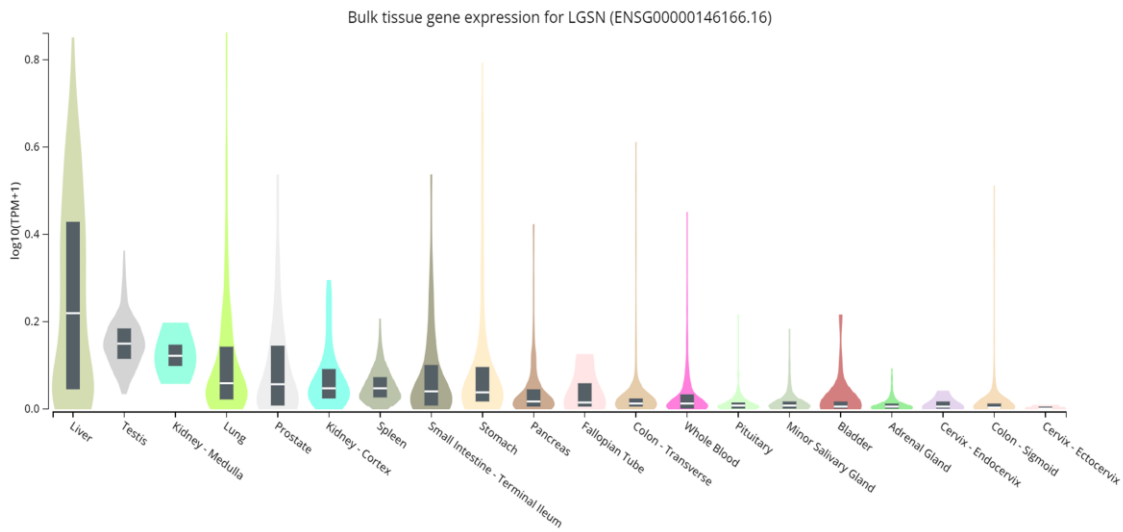

Figure 2A. Expression of Pathogenic variants influencing susceptibility genes for *Mycobacterium tuberculosis* infection in multiple continents for LGSN gene

Figure 2B. Expression of Pathogenic variants influencinge susceptibility genes for *Mycobacterium tuberculosis* infection in multiple continents for HLA-DQA1 gene

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

= 0.212). In contrast, rs41553512 showed only a weak correlation (r = 0.22, p = 0.778). While none proved to be statistically significant, likely due to the small number of regional data points, these associations highlight genetic variants of potential relevance to TB epidemiology. The results of the correlation analysis showed a significant association between the frequency of certain genetic variants (SNPs) and the incidence of tuberculosis (TB) in various world populations. Variants rs35269992, rs9272785, rs41533512, and rs2260497 showed a positive association with TB incidence, although the strength of the correlation varied, with rs9272785 showing the strongest association (Figure 4). This indicates that these genetic variants may play a role in TB susceptibility, especially through their influence on the body's immune response, such as cytokine production and macrophage activity. In the context of pharmacogenomics, identification of these variants may help develop more targeted TB therapies, including immunotherapy and personalised treatment based on the patient's genetic profile. These findings highlight opportunities for further research into the specific role of these SNPs in TB pathogenesis and the development of genetic-based therapies.

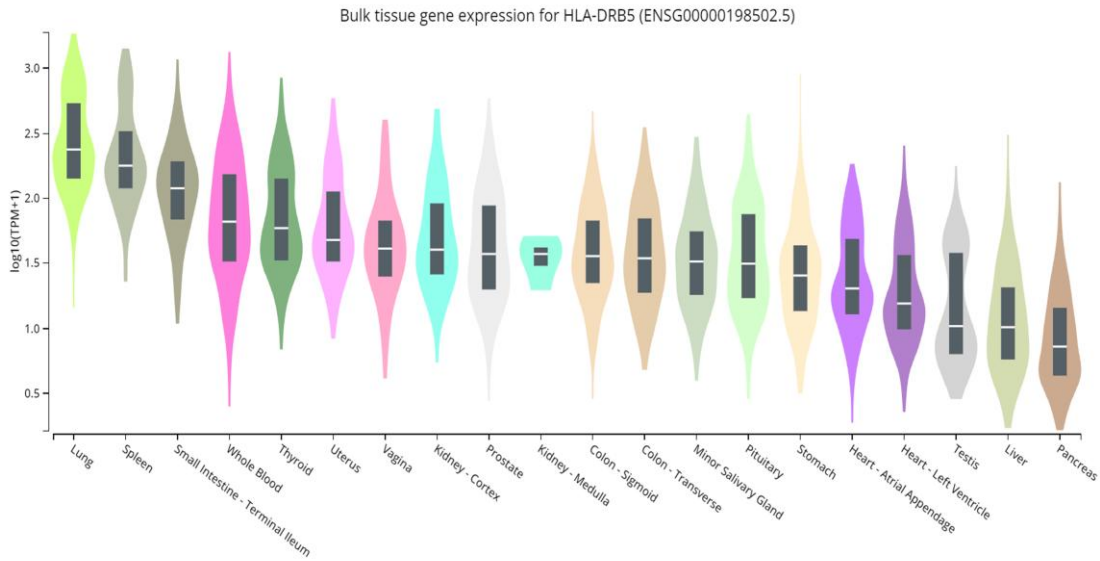

Figure 2C. Expression of pathogenic variants influencing susceptibility genes for *Mycobacterium tuberculosis* infection in multiple continents for HLA-DRB5 gene

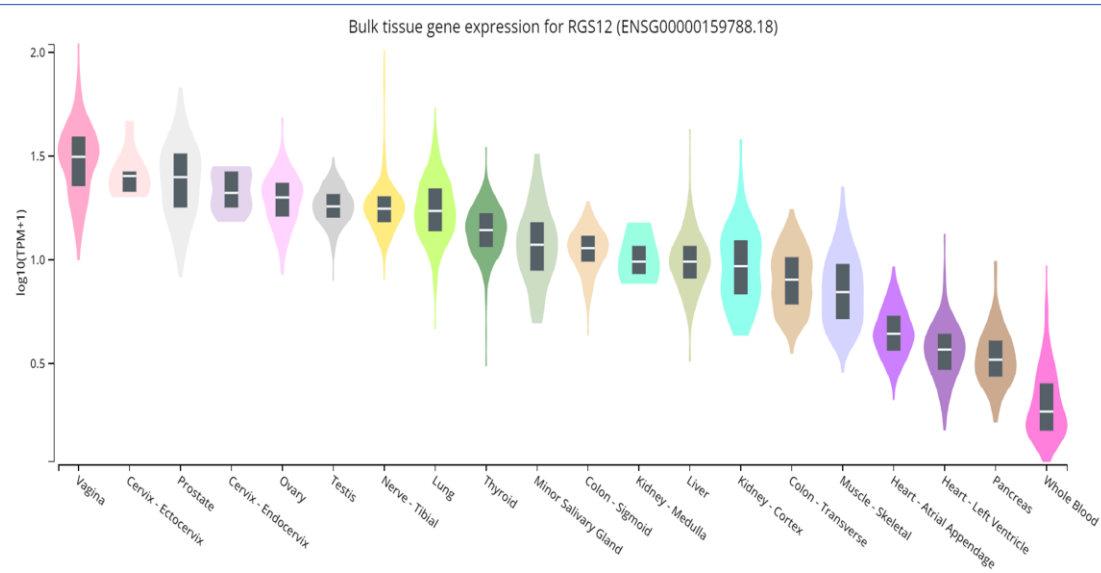

Figure 2D. Expression of Pathogenic variants influencing susceptibility genes for *Mycobacterium tuberculosis* infection in multiple continents for RGS12 gene

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

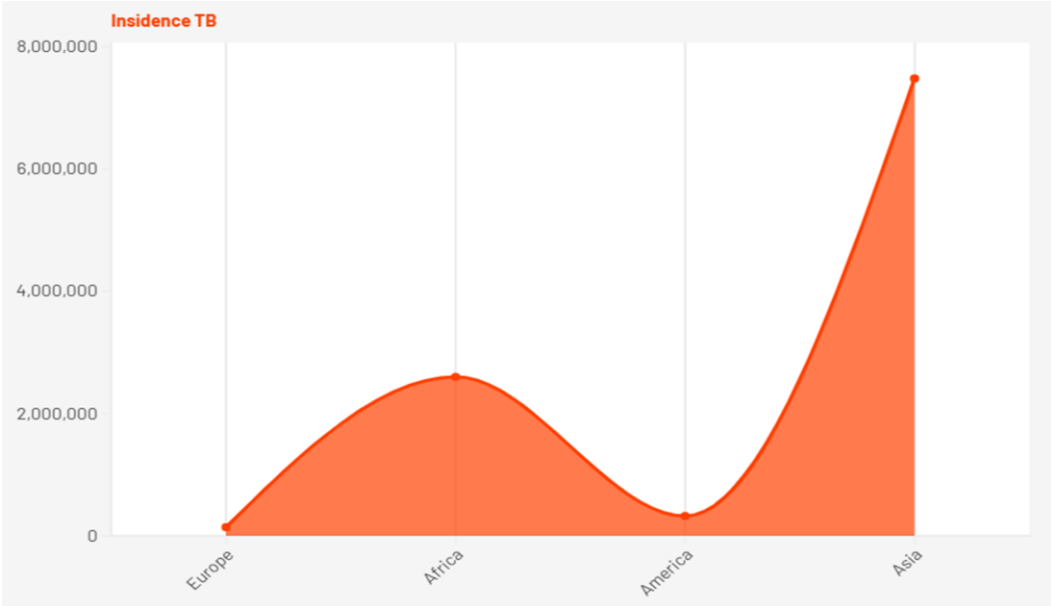

Figure 3. Number of incident cases of TB disease in the world

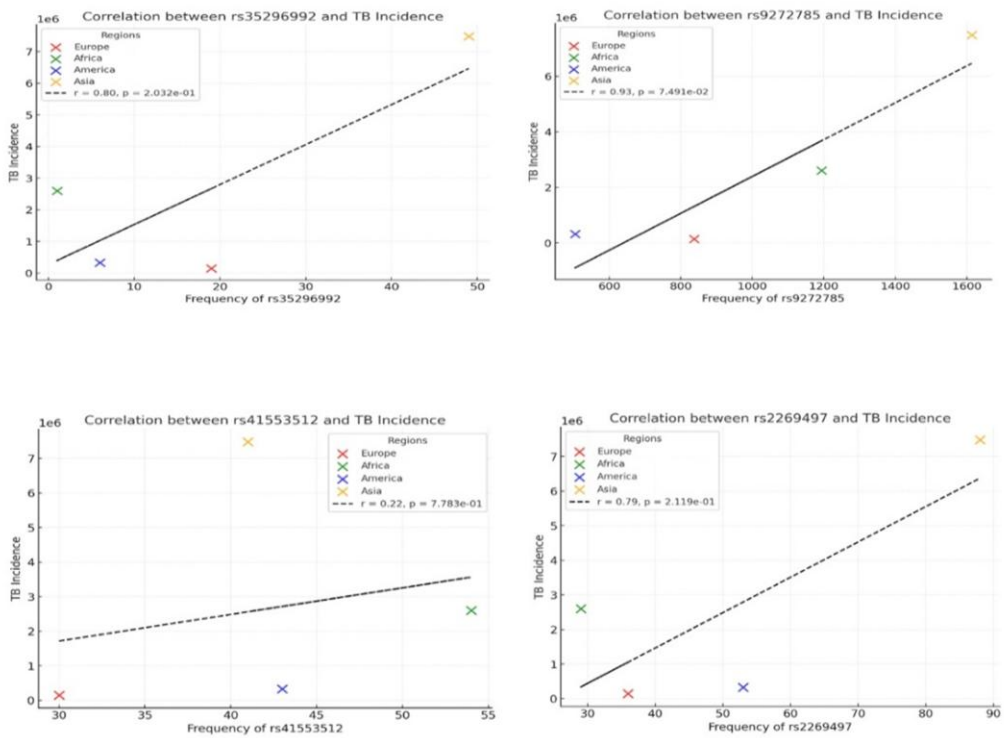

Figure 4. Correlation of allele frequency of variants rs35296992, rs9272785, rs41553512, rs2269497 with the incidence of TB disease in the world. Correlation of allele frequency of variants (rs35296992, rs9272785, rs41553512, rs2269497) with the incidence of tuberculosis (TB) across global regions (Europe, Africa, America, and Asia). Pearson's correlation coefficient (r) and corresponding p-values are presented in each panel. A positive correlation was observed for rs35296992 (r = 0.80, p = 0.203), rs9272785 (r = 0.93, p = 0.075), and rs2269497 (r = 0.79, p = 0.212), while rs41553512 showed a weak correlation with TB incidence (r = 0.22, p = 0.778). Although not statistically significant due to limited sample size, these trends suggest potential genetic contributions to TB susceptibility.

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

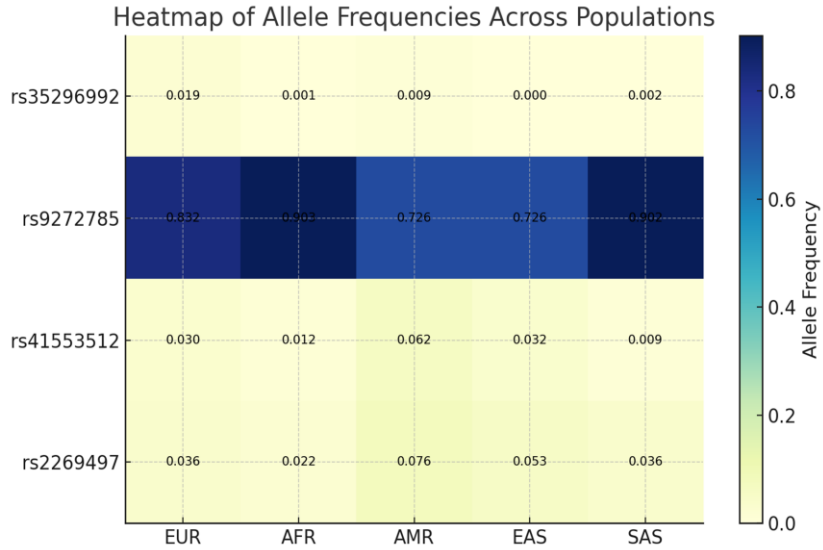

Figure 5. Heatmap of allele frequencies of selected SNPs across global populations. The heatmap displays allele frequency distributions for four SNPs (rs35296992 in LGSN, rs9272785 in HLA-DQA1, rs41553512 in HLA-DRB5, and rs2269497 in RGS12) across five major populations: European (EUR), African (AFR), Admixed American (AMR), East Asian (EAS), and South Asian (SAS). Darker blue shades indicate higher allele frequencies, while lighter shades indicate lower frequencies. Numerical values within each cell represent the exact allele frequency for the respective SNP and population

Table 1. Variant associated tuberculosis with criterion p-value <10-5

SNP	p-value	chr_pos
rs35296992	1,00E-07	chr6: 63990648
rs9272785	6,00E-06	chr6:32610401
rs9272785	6,00E-06	chr6: 32486402
rs41553512	8,00E-11	chr4: 3429856
rs2269497	3,00E-08	chr6: 63990648

Table 2. Allele frequencies for Pathogenic variants that influence susceptibility genes for *Mycobacterium tuberculosis* infection in multiple continents

SNP	Position (hg19) (bp)	Gene	Functional annotation	Amino Acid Change	Allele		Allele Frequencies of population genetics (N)				
					Ref	Eff*	EUR (503)	AFR (661)-	AMR (347)	EAS (504)	SAS (489)
rs35296992	chr6: 63990648	*LGSN*	Missense	p.Phe270Leu	A	G	0.019 (19)	0.001 (1)	0.009 (6)	0.000 (2)	0.002 (47)
rs9272785	chr6:32610401	*HLA-DQA1*	Missense	p.Ala210Thr	A	G	0.832 (837)	0.903 (1194)	0.726 (504)	0.726 (732)	0.902 (882)
rs41553512	chr6: 32486402	*HLA-DRB5*	Missense	p.Val232Ile	C	T	0.030 (30)	0.012 (54)	0.062 (43)	0.032 (32)	0.009 (9)
rs2269497	chr4: 3429856	*RGS12*	Missense	p.Asn1124Ser	A	G	0.036 (36)	0.022 (29)	0.076 (53)	0.053 (53)	0.036 (35)

* EUR, European; AFR, African; AMR, American; EAS, East Asian; SAS, Southeast Asian. N: total number of samples, Ref, Reference and Eff, Effect allele of EUR, AFR, AMR, EAS, SAS, Ensembl,org (http://www.ensembl.org/Homo_sapiens/Variation). * Effect allele is defined as the allele associated with higher expression for Pathogenic variants that influence susceptibility genes for *Mycobacterium tuberculosis* infection.

DISCUSSION

The aetiology of tuberculosis is multifaceted, involving a complex interplay between environmental exposures, host genetics, immunological status, and behavioural factors. Addressing these diverse influences is crucial to developing effective prevention strategies and improving outcomes for individuals at risk of TB infection across diverse populations worldwide [33]. Understanding these factors can inform targeted public health interventions to reduce the burden of tuberculosis among vulnerable groups. The article provides epidemiological data on the susceptibility to *Mycobacterium tuberculosis* infection, which varies significantly across populations and geographical regions, using genomic variants. Our study identified four unique variants with missense mutations associated with *Mycobacterium tuberculosis* infection across multiple continents: rs35296992, which encodes LGSN gene;

rs9272785, which encodes HLA-DQA1 gene; rs41553512, which encodes HLA-DRB5 gene; and rs2269497, which encodes RGS12 gene. We evaluated the distribution of allele frequencies in various populations, including European, African, American, East Asian (EAS), and Southeast Asian groups. Additionally, these genes were found to be expressed not only in the liver and lungs but also in the blood. Understanding the genetic basis of susceptibility to M. tuberculosis infection is critical for developing targeted interventions and improving disease management strategies across diverse populations worldwide. Continued research into these susceptibility genes will enhance our knowledge of host-pathogen interactions and may lead to novel therapeutic approaches for TB prevention and treatment.

We analysed the distribution of allele frequencies for the aforementioned genetic variants across various populations, specifically European, African, American, East Asian (EAS), and Southeast Asian groups. The findings revealed that the expression of these genes is not confined to the liver and lungs; they are also present in the blood. The study emphasises that these four specific variants are associated with susceptibility to M. tuberculosis infection, indicating their role in the immune response [34]. By encompassing multiple continents, the research highlights the global significance of these genetic variants in understanding tuberculosis susceptibility. Population-specific allele frequency patterns emerged across the five 1000 Genomes superpopulations. rs35296992 (LGSN) showed the highest frequency in Europeans, followed by Admixed Americans, South Asians, and East Asians. In contrast, rs9272785 (HLA-DQA1) was most prevalent in Africans, with decreasing frequencies observed in South Asians, Europeans, East Asians, and Admixed Americans. rs41553512 (HLA-DRB5) and rs2269497 (RGS12) both demonstrated the highest frequencies in Admixed Americans, followed by East Asians, Europeans, South Asians, and Africans. These distinct distribution patterns underscore population-specific genetic risk profiles for TB susceptibility.

The variant rs41553512 showed the highest distribution in Americans, and its prevalence decreased in East Asians, Europeans, Africans, and Southeast Asians, suggesting that genetic factors influencing TB risk may vary significantly across these regions. rs2269497 and rs41553512 were also predominantly found in Americans, followed by East Asians, Europeans, Southeast Asians, and Africans. This pattern reinforces the notion that certain genetic variants may confer varying levels of TB risk depending on geographic and ethnic backgrounds. According to Table 2, Americans exhibited a notably higher distribution of SNPs for two variants: rs41553512 and rs2269497. This observation suggests that these SNPs may play a crucial role in understanding TB susceptibility within American populations. Conversely, the highest distribution for Africans was observed with rs9272785, indicating a unique genetic risk profile for TB that differs from other populations. Furthermore, Europeans displayed the highest distribution for rs35296992, which underscores the importance of considering population-specific genetic factors when studying disease susceptibility. Overall, these findings underscore the complexity of TB susceptibility as influenced by genetic variants across different populations. Understanding these distributions can inform targeted prevention strategies and therapeutic interventions for specific demographic groups [35,36].

Additionally, the assessment of allele frequencies across different populations suggests that variations in susceptibility may exist based on genetic backgrounds. The expression of these identified genes in lung, liver, and whole blood implies important functional roles in immune responses to infection [37]. Mechanistically, some of these variants may influence TB pathogenesis through their involvement in critical immune processes. For example, the variants in HLA-DQA1 (rs9272785) and HLA-DRB5 (rs41553512) genes likely affect antigen presentation, which is essential for initiating adaptive immune responses against *Mycobacterium tuberculosis*. Altered antigen presentation may impact T-cell activation and cytokine signalling pathways, thereby modulating the host immune response to infection. Similarly, variants in RGS12 (rs2269497), which regulates signalling pathways, could influence cytokine-mediated inflammatory responses, affecting disease progression and severity. Although less characterised, the LGSN variant (rs35296992) may impact immune or metabolic pathways in the liver that contribute systemically to host defence mechanisms in tuberculosis [PMID: 32033104], [PMID: 38812605],[PMID: 30167433]. This information contributes to a broader understanding of how genetic factors influence susceptibility to tuberculosis across diverse populations and may guide future research and therapeutic strategies. Furthermore, advancements in genomic techniques, such as next-generation sequencing (NGS) and RNA sequencing, have identified numerous single-nucleotide polymorphisms (SNPs), insertions and deletions (indels), rearrangements, gene fusions, and copy number variations across different malignancies, with the aim of identifying druggable targets for treatment [38,39].

The term “non-synonymous changes” refers to alterations in the DNA that lead to a different amino acid being produced compared to the original sequence [28]. This is contrasted with “synonymous changes,” in which the amino acid remains the same despite a DNA change. In this analysis, four specific genes were identified that contain missense mutations. These are LGSN: The mutation involves a change from phenylalanine (p.Phe270) to leucine (Leu) at position 270 of the protein. HLA-DQA1: This variant results in a change from alanine (p.Ala210) to threonine (Thr) at position 210. HLA-DRB5: A change from valine (p.Val232) to isoleucine (Ile) at position 232. RGS12: This mutation involves a substitution of asparagine (p.Asn1124) with serine (Ser) at position 1124. Each of these changes can potentially alter the structure and

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

function of the resulting protein. The specific amino acids involved can affect protein folding, stability, and interactions with other molecules, which may have implications for biological processes such as the immune response. Identifying these missense mutations is important for understanding how genetic variations may influence susceptibility to diseases, such as tuberculosis, by affecting protein function. Our findings outline a systematic approach to mapping genetic variants onto genes, highlighting specific missense mutations that could impact protein function and contribute to understanding disease susceptibility [39].

Based on gene expression in various tissues, our analysis revealed that the four genes—rs35296992 (encoding LGSN), rs9272785 (encoding HLA-DQA1), rs41553512 (encoding HLA-DRB5), and rs2269497 (encoding RGS12)—show distinct expression patterns in lung, liver, and whole blood. Understanding these expression differences is crucial for elucidating their roles in tuberculosis (TB) pathology. *LGSN* (rs35296992). This gene showed significantly higher expression in the liver than in either the lung or whole blood. The elevated expression in the liver may suggest a role in systemic immune responses or metabolic processes that could influence TB susceptibility or progression. In contrast, the *HLA-DQA1* gene exhibited the highest expression in lung tissue relative to other tissues. This finding is particularly relevant given the lungs' primary role as the site of *Mycobacterium tuberculosis* infection. The high expression of HLA-DQA1 in lung tissue likely enhances MHC class II antigen presentation to CD4+ T cells, facilitating activation of adaptive immune responses critical for TB containment and control. Similar to *HLA-DQA1*, *HLA-DRB5* also demonstrated higher expression in the lungs, followed by whole blood and liver. This pattern suggests that *HLA-DRB5* may play a vital role in pulmonary immune responses, potentially influencing human host response to TB infection. The RGS12 gene also showed higher expression in lung tissue than in liver or whole blood. This finding indicates that RGS12 may be involved in signalling pathways relevant to lung immunity or inflammation, which could be significant for understanding TB pathogenesis.

These findings underscore the importance of identifying pathogenic variants that influence differential gene expression across various tissues relevant to TB pathology. Understanding how these genes are expressed differently can provide insights into their functional roles in TB susceptibility and disease progression. For instance, genes highly expressed in lung tissue may be crucial for local immune responses against *Mycobacterium tuberculosis*, while those with higher expression in the liver might be involved in systemic responses or metabolic regulation. However, it is important to note that our analysis used bulk tissue expression data from GTEx, which represents averaged signals from heterogeneous cell populations. This approach may mask cell-type-specific expression patterns that could provide deeper insights into gene functions in TB pathology. This bulk tissue analysis may mask cell-type-specific expression patterns that could provide deeper insights into TB-relevant gene functions within immune cell populations. Single-cell RNA sequencing data would enable discrimination of these cell-specific expression profiles; however, such high-resolution datasets remain limited for human lung and TB-affected tissues [40,41]. Therefore, our findings should be interpreted with regard to cell-specific contributions. Overall, the differential expression of these genes highlights the complexity of host-pathogen interactions in TB and underscores the need for further research into how genetic variants influence gene expression across tissues. Such understanding can inform targeted therapeutic strategies and improve our grasp of individual susceptibility to TB based on genetic backgrounds and tissue-specific immune responses.

Importantly, *HLA* variants, especially those in *HLA-DQA1* and *HLA-DRB5*, play critical roles in TB pathogenesis. These variants affect antigen presentation to CD4+ T cells, which are essential for initiating and modulating adaptive immune responses against M. tuberculosis. Polymorphisms in HLA class II molecules influence peptide-binding affinity and the repertoire of presented antigens, thereby modulating cytokine profiles and T-cell activation, which are essential for effective infection control. Additionally, emerging evidence shows that certain HLA alleles correlate with altered M. tuberculosis–specific T cell responses and distinct gene expression patterns, highlighting their functional significance in host defence. Such immunogenetic diversity across populations partially explains global differences in TB susceptibility and informs the design of vaccines and immunotherapies aimed at broad efficacy across diverse HLA backgrounds. The translational implications of our findings are significant. By identifying genetic variants that influence immune responses to M. tuberculosis, our study provides preliminary data for personalised approaches to TB control. In vaccine design, understanding HLA variant distributions can guide the selection of antigenic epitopes that broadly cover diverse HLA types, thereby enhancing vaccine efficacy across different populations [42,43]. Moreover, identifying variants that modulate immune signalling pathways, such as those in RGS12, may inform the development of host-directed therapies to boost protective immune mechanisms or mitigate harmful inflammation.

Furthermore, genetic risk stratification based on these variants could enable targeted screening and preventive interventions for high-risk individuals, improving early detection and clinical outcomes [44–46]. These insights align with precision medicine goals regarding targeted TB prevention and treatment strategies according to genetic susceptibility profiles, ultimately contributing to more effective and equitable TB control globally. We acknowledge the complexity of correlating allele frequencies with TB incidence, as environmental factors (e.g., malnutrition, HIV coinfection), socioeconomic status,

share
Send us an email: hsijournal@ug.edu.gh
Visit us: https://www.hsijournal.ug.edu.gh

healthcare access, and pathogen diversity also contribute significantly to TB burden. These confounders may obscure direct genetic effects and necessitate multifactorial analytical approaches. Nonetheless, consistent associations across populations provide compelling leads for functional investigation and translational advances. The translational potential of our findings is substantial. Understanding HLA-mediated immune responses enables refined vaccine design strategies that incorporate epitopes with broad binding across prevalent HLA variants or target less polymorphic molecules, such as HLA-E, to elicit robust T cell immunity. Moreover, identifying pathogenic variants can inform personalised host-directed therapies that modulate immune pathways, enhance macrophage antimicrobial activity, or counteract immune evasion mechanisms in genetically predisposed individuals. Genetic screening for risk alleles may also facilitate targeted prophylaxis and early therapeutic interventions, aligning with precision medicine goals to optimise TB control in high-risk populations.

Our findings indicate that several variants identified in this study have not been previously reported, based on bioinformatics and functional annotations. However, we acknowledge certain limitations inherent to our current study; first, not all variants are pathogenic. Consequently, some of the variants associated with TB risk may not correlate with gene expression. This highlights the importance of thorough functional annotation in bioinformatics, which is essential for understanding the biological significance of genomic variations and their potential implications in disease contexts. Importantly, our study lacks experimental validation, such as functional assays, to confirm the biological effects of the prioritised variants and their mechanistic roles in TB susceptibility. To address this, future studies could employ CRISPR-Cas9 knock-in models in relevant human cell lines or animal models to introduce specific missense mutations and assess their functional impact on immune responses and infection outcomes. Additionally, expression quantitative trait loci (eQTL) mapping in patient-derived samples could help link genetic variants to gene expression changes in key tissues. Complementary immunological assays, such as antigen presentation and T-cell activation assays for HLA gene variants, or cytokine signalling assays for RGS12 variants, would further elucidate mechanisms involved.

Experimental validation is critical for substantiating bioinformatics predictions and bridging the gap to clinical applications. Additionally, potential bias in Genome-wide association study (GWAS) data due to underrepresentation of certain populations, particularly those from low-resource settings or indigenous groups disproportionately affected by tuberculosis, may limit the generalizability and inclusiveness of our findings. Expanding GWAS datasets to encompass diverse populations remains an important future direction to ensure an equitable understanding of genetic susceptibility worldwide. Functional annotation helps researchers interpret large-scale omics data and identify genes associated with specific traits or diseases, facilitating biomedical research and drug discovery. However, challenges remain, including the need to determine the draggability of identified variants and their relevance to gene expression profiles. While this study provides new insights into TB-related variants, further research, including experimental validation and more diverse population sampling, is necessary to fully explore their functional implications and clinical relevance.

Conclusion

Our findings emphasise the significant role of specific SNPs rs35296992 (*LGSN*), rs9272785 (*HLA-DQA1*), rs41553512 (*HLA-DRB5*), and rs2269497 (*RGS12*) in influencing susceptibility to tuberculosis (TB). Correlation tests showed that these variants were positively associated with TB incidence in various populations, with rs9272785 showing the strongest correlation, followed by rs35296992. Although the variants rs41553512 and rs2269497 showed weaker correlations, both still contributed to TB risk. Differentiation of gene expression across tissues, such as the lung, liver, and blood, is essential to understanding how these genes contribute to TB pathogenesis. Knowledge of the tissue-specific functions of these genes may deepen our understanding of their contribution to the immune response and to susceptibility to TB, thereby informing the development of more targeted therapeutic strategies for the management of this disease.

DECLARATIONS

Ethical consideration

Ethical approval was not required for this study, as it involved solely the analysis of data sourced from publicly accessible databases. No direct involvement with human subjects or collection of new personal information took place. The datasets used were previously collected in accordance with appropriate ethical standards by the original data custodians.

Consent to publish

All authors agreed on the content of the final paper.

Funding

This study is supported by the Lembaga Penelitian dan Pengabdian kepada Masyarakat of Universitas Ahmad Dahlan, under agreement numbers PD-229/SP3/LPPM-UAD/XI/2024 and U12/1529/XI/2024.

Competing Interest

The authors declare no conflict of interest

Author contribution

LMI and WA conceived and designed the study. LMI, RC, and SK performed the computational analysis. LMI drafted the manuscript. LMI, WA, SK, INS, IH, MI, IA, SP, DR, BHR, and RC reviewed and revised the manuscript. LMI supervised and coordinated the study. All authors read and approved the final

manuscript and made significant contributions to the study.

Acknowledgement
We would like to express our sincere gratitude to the Bidang Riset dan Inovasi and the entire team at Universitas Ahmad Dahlan for their invaluable support throughout this research. We dedicate this work to all individuals affected by tuberculosis in our country, encouraging them to remain resilient and hopeful. Although diseases cannot always be prevented, we as humans can strive to adopt the best possible preventive measures. Through collaboration with our esteemed colleagues, we aim to contribute valuable knowledge and effective solutions to combat tuberculosis, with the shared aspiration of achieving zero cases in the near future.

Availability of data
Data is available upon request to the corresponding author

REFERENCES

1. World Health Organization (2021) Tuberculosis.
2. Falzon D, Zignol M, Bastard M, Floyd K, Kasaeva T (2023) The impact of the COVID-19 pandemic on the global tuberculosis epidemic. Front Immunol. 14:1234785
3. Narasimhan P, Wood J, Macintyre CR, Mathai D (2013) Risk factors for tuberculosis. Pulm Med. 2013:828939
4. Beiranvand R, Karimi A, Delpisheh A, Sayehmiri K, Soleimani S, Ghalavandi S (2016) Correlation assessment of climate and geographic distribution of tuberculosis using geographical information system (GIS). Iran J Public Health. 45:86–93
5. Lei Y, Wang J, Wang Y, Xu C (2023) Geographical evolutionary pathway of global tuberculosis incidence trends. BMC Public Health. 23:755
6. Zheng R, Li Z, He F, Liu H, Chen J, Chen J, Xie X, Zhou J, Chen H, Wu X, et al. (2018) Genome-wide association study identifies two risk loci for tuberculosis in Han Chinese. Nat Commun. 9:4072
7. Thye T, Vannberg FO, Wong SH, Owusu-Dabo E, Osei I, Gyapong J, Sirugo G, Sisay-Joof F, Enimil A, Chinbuah MA, et al. (2010) Genome-wide association analyses identifies a susceptibility locus for tuberculosis on chromosome 18q11.2. Nat Genet. 42:739–741
8. He C, Cheng X, Kaisaier A, Wan J, Luo S, Ren J, Sha Y, Peng H, Zhen Y, Liu W, et al. (2022) Effects of Mycobacterium tuberculosis lineages and regions of difference (RD) virulence gene variation on tuberculosis recurrence. Ann Transl Med. 10:49
9. Uren C, Henn BM, Franke A, Wittig M, van Helden PD, Hoal EG, Möller M (2017) A post-GWAS analysis of predicted regulatory variants and tuberculosis susceptibility. PLoS One. 12:e0174738
10. Sun W, Jiao L, Liu T, Song J, Wang M, Liang L, Wen C, Hu L, Qu W, Ying B (2020) No significant effects of IL-6 and IL-13 gene variants on tuberculosis susceptibility in the Chinese population. DNA Cell Biol. 39:1356–1367
11. McHenry ML, Simmons J, Hong H, Malone LL, Mayanja-Kizza H, Bush WS, Boom WH, Hawn TR, Williams SM, Stein CM (2023) Tuberculosis severity associates with variants and eQTLs related to vascular biology and infection-induced inflammation. PLoS Genet. 19:e1010387
12. Correa-Macedo W, Cambri G, Schurr E (2019) The interplay of human and Mycobacterium tuberculosis genomic variability. Front Genet. 10:865
13. Xu M, Li J, Xiao Z, Lou J, Pan X, Ma Y (2020) Integrative genomics analysis identifies promising SNPs and genes implicated in tuberculosis risk based on multiple omics datasets. Aging (Albany NY). 12:19173–19220
14. Soha A, Azina I, Rozentale B, Kramicha K, Sture G, Savicka O, Titovica G (2024) HLA class II DRB1, DQA1, DQB1 loci in patients with HIV infection and tuberculosis in a Latvian cohort group. Cent Eur J Immunol. 49:37–44
15. Llibre A, Smith N, Rouilly V, Musvosvi M, Nemes E, Posseme C, Mabwe S, Charbit B, Mbandi SK, Filander E, Africa H, Saint-André V, Bondet V, Bost P, Mulenga H, Bilek N, Albert ML, Scriba TJ, Duffy D (2022) Tuberculosis alters immune-metabolic pathways resulting in perturbed IL-1 responses. Front Immunol. 13:897193
16. Faksri K, Tan JH, Chaiprasert A, Teo YY, Ong RT (2016) Bioinformatics tools and databases for whole genome sequence analysis of Mycobacterium tuberculosis. Infect Genet Evol. 45:359–368
17. Cuevas-Córdoba B, Fresno C, Haase-Hernández JI, Barbosa-Amezcua M, Mata-Rocha M, Muñoz-Torrico M, Salazar-Lezama MA, Martínez-Orozco JA, Narváez-Díaz LA, Salas-Hernández J, et al. (2021) A bioinformatics pipeline for Mycobacterium tuberculosis sequencing that cleans contaminant reads from sputum samples. PLoS One. 16:e0258774
18. Amukti DP, Irham LM, Surono S, Adikusuma W, Chong R, El Khair R, et al. (2024) Genomic variants and epidemiology of atherosclerosis—worldwide correlation analysis and utilisation of atherosclerosis gene variants for identification of drug target candidates with bioinformatics approach. Bulg Cardiol. 30:108–120
19. Humolungo DT, Anjani R, Irham LM, Sulistyani N, Ma'ruf M, Amukti DP, Adikusuma W, Sarasmita MA, Khairi S, Purwanto BD, Suyatmi S (2024) Identification of pathogenic gene variants in carpal tunnel syndrome using bioinformatics approaches. In: E3S Web Conf. 501:01022. EDP Sciences
20. Amukti DP, Wazni AR, Irham LM, Sulistyani N, Ma'ruf M, Adikusuma W, Sarasmita MA, Khairi S, Purwanto BD, Suyatmi S, Siswanto LM (2024) Identifying pathogenic variants associated with Alzheimer by integrating genomic databases and bioinformatics approaches. In: E3S Web Conf. 501:01021. EDP Sciences
21. Gumelar G, Ulfa MM, Amukti DP, Irham LM, Yuliani S, Adikusuma W, et al. (2024) Harnessing genomic and bioinformatic data to broaden understanding of leukaemia across continents. Scripta Med (Banja Luka). 55:717–725
22. Hubbard T, Barker D, Birney E, Cameron G, Chen Y, Clark L, Cox T, Cuff J, Curwen V, Down T, et al. (2002) The Ensembl genome database project. Nucleic Acids Res. 30:38–41
23. Aken BL, Ayling S, Barrell D, Clarke L, Curwen V, Fairley S, Fernandez Banet J, Billis K, García Girón C, Hourlier T,

et al. (2016) The Ensembl gene annotation system. Database (Oxford). 2016:baw093

24. Made I, Wijaya H, Fahrian Hipmi A, Nur S, Abas R, Amukti DP, et al. (2024) Potensi variasi gen FGL1 dalam peningkatan risiko efek samping pada hati akibat penggunaan nifedipin pada pasien hipertensi gestasional. J Ilm Farm Farm Klin (JIFFK). 21:269–276

25. Irham LM, Adikusuma W, Lolita L, Puspitaningrum AN, Afief AR, Sarasmita MA, Dania H, Khairi S, Djalilah GN, Purwanto BD, et al. (2023) Investigation of susceptibility genes for chickenpox disease across multiple continents. Biochem Biophys Rep. 33:101419

26. Pakha DN, Yudhani RD, Irham LM (2024) Investigation of missense mutation-related type 1 diabetes mellitus through integrating genomic databases and bioinformatic approach. Genomics Inform. 22:8

27. Ma'ruf M, Fadli JC, Mahendra MR, Irham LM, Sulistyani N, Adikusuma W, Chong R, Septama AW (2023) A bioinformatic approach to identify pathogenic variants for Stevens-Johnson syndrome. Genomics Inform. 21:e26

28. Chu D, Wei L (2019) Non-synonymous, synonymous and nonsense mutations in human cancer-related genes undergo stronger purifying selections than expectation. BMC Cancer. 19:359

29. Shahzad F, Bashir N, Ali A, Nadeem A, Ammar A, Kashif M, Javaid K, Jahan S, Tahir R, Rizwan M, Mushtaq A, Afzal N (2022) SLC11A1 genetic variation and low expression may cause immune response impairment in TB patients. Genes Immun. 23:85–92

30. Emadi E, Akhoundi F, Kalantar SM, Emadi-Baygi M (2020) Predicting the most deleterious missense nsSNPs of the protein isoforms of the human HLA-G gene and in silico evaluation of their structural and functional consequences. BMC Genet. 21:94

31. Scott RJ, Meldrum CJ (2005) Missense mutations in cancer predisposing genes: can we make sense of them? Hered Cancer Clin Pract. 3:123–127

32. Prasetyaning Amukti D, Pratami RI, Farida Ismyama D, Puspitasari A, Herlina T (n.d.) Pharmacogenomic and bioinformatic insights into ACE gene variants and their influence on ACE inhibitor response in hypertension. Pharm Clin J Nusantara. Available from: https://nusantarascientificjournal.com/index.php/pcjn/index

33. Vasiliu A, Martinez L, Gupta RK, Hamada Y, Ness T, Kay A, Bonnet M, Sester M, Kaufmann SHE, Lange C, et al. (2024) Tuberculosis prevention: current strategies and future directions. Clin Microbiol Infect. 30:1123–1130

34. Satta G, Witney AA, Shorten RJ, Karlikowska M, Lipman M, McHugh TD (2016) Genetic variation in Mycobacterium tuberculosis isolates from a London outbreak associated with isoniazid resistance. BMC Med. 14:117

35. Aziz R, Sen P, Beura PK, Das S, Tula D, Dash M, Namsa ND, Deka RC, Feil EJ, Satapathy SS, et al. (2022) Incorporation of transition to transversion ratio and nonsense mutations improves the estimation of the number of synonymous and non-synonymous sites in codons. DNA Res. 29:dsac023

36. Lee CW, Kim SM, Sa S, Hong M, Nam SM, Han HW (2023) Relationship between drug targets and drug-signature networks: a network-based genome-wide landscape. BMC Med Genomics. 16:17

37. Tarek M, Abdelzaher H, Kobeissy F, El-Fawal HAN, Salama MM, Abdelnaser A (2021) Bioinformatics analysis of allele frequencies and expression patterns of ACE2, TMPRSS2 and FURIN in different populations and susceptibility to SARS-CoV-2. Genes (Basel). 12:1041

38. Suliman S, Nieto-Caballero VE, Asgari S, Lopez K, Iwany SK, Luo Y, Nathan A, Fernandez-Salinas D, Chiñas M, Huang CC, et al. (2023) History of tuberculosis disease is associated with genetic regulatory variation in Peruvians. medRxiv. 10.1101/2023.06.20.23291558

39. Advani J, Verma R, Chatterjee O, Pachouri PK, Upadhyay P, Singh R, Yadav J, Naaz F, Ravikumar R, Buggi S, et al. (2019) Whole genome sequencing of Mycobacterium tuberculosis clinical isolates from India reveals genetic heterogeneity and region-specific variations that might affect drug susceptibility. Front Microbiol. 10:309

40. Donovan MKR, D'Antonio-Chronowska A, D'Antonio M, Frazer KA (2020) Author correction: cellular deconvolution of GTEx tissues powers discovery of disease and cell-type associated regulatory variants. Nat Commun. 11:4426

41. Kim-Hellmuth S, Aguet F, Oliva M, Muñoz-Aguirre M, Kasela S, Wucher V, et al. (2020) Cell type–specific genetic regulation of gene expression across human tissues. Science. 369:6509

42. Mahmood MS, Bin-T-Abid D, Irshad S, Batool H (2021) Analysis of putative epitope candidates of Mycobacterium tuberculosis against Pakistani human leukocyte antigen background: an immunoinformatic study for the development of future vaccine. Int J Pept Res Ther. 27:597–614

43. Shiraz Mohd, Lata S, Kumar P, Shankar UN, Akif Mohd (2021) Immunoinformatics analysis of antigenic epitopes and designing of a multi-epitope peptide vaccine from putative nitro-reductases of Mycobacterium tuberculosis DosR. Infect Genet Evol. 94:105017

44. Andongma BT, Huang Y, Chen F, Tang Q, Yang M, Chou SH, et al. (2023) In silico design of a promiscuous chimeric multi-epitope vaccine against Mycobacterium tuberculosis. Comput Struct Biotechnol J. 21:991–1004

45. Albutti A (2021) An integrated computational framework to design a multi-epitopes vaccine against Mycobacterium tuberculosis. Sci Rep. 11:21929

46. Nayak SS, Sethi G, Ramadas K (2023) Design of multi-epitope based vaccine against Mycobacterium tuberculosis: a subtractive proteomics and reverse vaccinology based immunoinformatics approach. J Biomol Struct Dyn. 41:14116–14134