J Korean Med Sci. 2026 Oct 26;41(41):e262
https://doi.org/10.3346/jkms.2026.41.e262
eISSN 1598-6357·pISSN 1011-8934

JKMS

Original Article
Respiratory Diseases & Critical Care Medicine

Check for updates

Investigation of Matrix Metalloproteinase–Related Genetic Variations Associated With Cavity Formation in Pulmonary Tuberculosis

Sungmin Zo,[1*] **Hayoung Choi**,[2*] **Sang Hyuk Kim**,[3] **Ye Eun Yoon**,[4] **Jae Won Yun**,[4] **and Hyun Lee**[5]

[1]Division of Respiratory and Critical Care Medicine, Department of Internal Medicine, Korea University Anam Hospital, Korea University College of Medicine, Seoul, Korea
[2]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, Hallym University Kangnam Sacred Heart Hospital, Hallym University College of Medicine, Seoul, Korea
[3]Division of Pulmonary, Allergy, and Critical Care Medicine, Department of Internal Medicine, Korea University Guro Hospital, Korea University College of Medicine, Seoul, Korea
[4]Veterans Health Service Medical Research Institute, Veterans Health Service Medical Center, Seoul, Korea
[5]Division of Pulmonary Medicine and Allergy, Department of Internal Medicine, Hanyang Medical Center, Hanyang University College of Medicine, Seoul, Korea

OPEN ACCESS

Received: Sep 17, 2025
Accepted: Feb 4, 2026
Published online: Jun 29, 2026

Address for Correspondence:
Jae Won Yun, MD, PhD
Veterans Health Service Medical Research Institute, Veterans Health Service Medical Center, 53 Jinhwangdo-ro 61-gil, Gangdong-gu, Seoul 05368, Korea.
Email: jwyunmd@gmail.com

Hyun Lee, MD, PhD
Division of Pulmonary Medicine and Allergy, Department of Internal Medicine, Hanyang Medical Center, Hanyang University College of Medicine, 222-1 Wangsimni-ro, Seongdong-gu, Seoul 04763, Korea.
Email: namuhanayeyo@hanyang.ac.kr

*Sungmin Zo and Hayoung Choi contributed equally to this work.

© 2026 The Korean Academy of Medical Sciences.
This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (https://creativecommons.org/licenses/by-nc/4.0/) which permits unrestricted non-commercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

ORCID iDs
Sungmin Zo
https://orcid.org/0000-0001-5993-8501
Hayoung Choi
https://orcid.org/0000-0003-4812-0653

ABSTRACT

Background: Matrix metalloproteinases (MMPs) play a key role in cavity formation in tuberculosis (TB) and are associated with disease progression and transmission. However, the genetic determinants of inter-individual variability, which influence the expression and signaling of MMP-related genes, remain poorly understood.
Methods: Small-scale whole-genome sequencing targeting MMPs was conducted on peripheral blood samples from 20 adult patients with TB to investigate genetic determinants of cavity formation. Following variant calling, statistical association analyses were performed using the Exact Association Test and Fisher's exact test as well as pathway and network analyses, with appropriate corrections for multiple testing.
Results: Eleven TB patients with cavitation and nine without cavitation were analyzed. Thirty genes related to extracellular matrix remodeling and MMP activity were identified as significantly associated with cavity formation using the Exact Association Test. Fisher's exact test revealed 35 single-nucleotide polymorphisms in patients with cavitary TB, including variants regulating key genes such as *KLKB1, LAMB3, LTBP4, ITGB2*, and *EGFR*. Network analysis demonstrated that the implicated genes were broadly distributed across the genome and linked to distinct biological pathways involved in TB cavitation.
Conclusion: This study identifies host genetic variants associated with TB cavity formation, particularly those involved in MMP activity, extracellular matrix remodeling, and immune signaling, supporting a multifactorial model of tissue destruction. These findings offer preliminary insights into host genetic contributors to cavitary TB.

Keywords: Pulmonary Tuberculosis; Cavity; Whole-Genome Sequencing

Generated by Xmlinkpress

Sang Hyuk Kim
https://orcid.org/0000-0002-0410-8524
Ye Eun Yoon
https://orcid.org/0009-0006-3872-9421
Jae Won Yun
https://orcid.org/0000-0002-9029-8036
Hyun Lee
https://orcid.org/0000-0002-1269-0913

Funding

This work was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIT) (RS-2025-00557268), the research fund of Hanyang University (HY-202500000001668), the Ministry of Education (grant No. RS-2025-25423084), and the Korea Health Industry Development Institute (KHIDI) funded by the Ministry of Health and Welfare, Republic of Korea (grant No. RS-2024-00408936). The funders had no role in the study design, data collection and analysis, or manuscript preparation.

Disclosure

Sang Hyuk Kim reports personal fees from AstraZeneca, Daewon, Handok, Hallym. He has received either in kind research support or funds paid to the institution from the Korea University Guro Hospital and Korean Academy of Tuberculosis and Respiratory Diseases.

Author Contributions

Conceptualization: Choi H, Lee H. Data curation: Kim SH. Formal analysis: Yoon YE, Yun JW. Investigation: Kim SH. Methodology: Zo S, Choi H. Supervision: Lee H. Writing – original draft: Zo S, Choi H. Writing – review & editing: Yun JW, Lee H.

INTRODUCTION

Tuberculosis (TB) is an infectious disease caused by airborne transmission of *Mycobacterium tuberculosis* and remains one of the leading causes of death worldwide.[1,2] As of 2023, Korea reported approximately 20,000 TB cases (38 cases per 100,000 people), ranking it second in incidence and fifth in mortality among Organization for Economic Co-operation and Development member countries. Given the country's transition into a super-aged society and the steadily increasing proportion of TB cases among individuals aged 65 and older, it is important to reduce TB-related mortality.

Cavity formation in TB plays a critical role in disease progression and transmission. Cavities serve as reservoirs for *M. tuberculosis*, hindering effective drug penetration, maintaining high bacterial loads, and increasing the risk of disease transmission.[3,4] Additionally, cavity formation has been associated with poor treatment outcomes, including greater risks of treatment failure and disease relapse.[4,5] Regarding the molecular pathogenesis of cavitary formation in TB, proteolytic enzymes—particularly matrix metalloproteinases (MMPs)—are major contributors by degrading the extracellular matrix (ECM) of lung tissue.[6-8] However, little is known about the genetic factors that promote differences among individuals in the expression and signaling of MMP-related genes.

Identifying host genetic factors associated with cavity formation is essential for advancing our understanding of TB pathogenesis and for optimizing individualized treatment strategies. While application of whole-genome sequencing (WGS) has enabled detection of single-nucleotide variants (SNVs), the complexity of next-generation sequencing data analysis continues to pose challenges for evaluating the comprehensive genetic background of individuals in TB research. Nevertheless, recent studies have suggested the feasibility of targeted or limited-scope next-generation sequencing approaches to investigate host genetic susceptibility in TB, particularly in the context of methodological constraints associated with WGS.[9,10] Building on this approach, the present study aimed to explore genetic variations targeting MMP genes that might contribute to differential susceptibility to cavity formation among TB patients.

METHODS

Study population and sequencing

This study included consecutive adult patients (aged ≥ 19 years) with newly diagnosed or retreated drug-susceptible cavitary TB at Hallym University Kangnam Sacred Heart Hospital (a 577-bed referral hospital in Seoul, Korea) between June 2019 and March 2021. Pulmonary TB was diagnosed based on sputum acid-fast bacilli smear and culture plus *M. tuberculosis* polymerase chain reaction results. The presence of cavities in study participants was determined by radiologists based on the chest radiographs recorded at the time of TB diagnosis.

At baseline, before anti-TB treatment was initiated, 19 mL of whole blood per patient was collected into ethylenediaminetetraacetic acid tubes. Within 4 hours of collection, samples were centrifuged at 480 g for 10 minutes at 4°C using an Eppendorf Centrifuge 5810 (No. 0012529-rotor A-4-81). WGS was performed at 30 × coverage using the NovaSeq 6000 platform (Illumina, San Diego, CA, USA). Library preparation was conducted

using the TruSeq DNA PCR-free kit (Illumina). After PCR duplicate removal using Picard MarkDuplicates, all samples achieved a mean sequencing depth of 30 ×–39 × (median 34 ×), ensuring consistently adequate coverage across samples for reliable variant detection.

Alignment, calling, and annotation

For genomic data analysis, sequence alignment was performed using the Isaac aligner (Illumina), mapping sequencing reads to the reference genome. Variants, including SNVs and small insertions/deletions (indels), were identified using the Isaac Variant Caller (Illumina),[11] and the results were generated in variant call format files. The detected variants were annotated using SnpEff,[12] with gene annotation based on hg19 genome coordinates. Additionally, the variants were mapped to dbSNP151[13] IDs and the 1000 Genomes Phase 1 Release v3 database[14] for reference.

Gene and pathways database utilization

To identify genes and pathways potentially associated with cavitation, data on MMP-related genes and pathways were obtained from the ConsensusPathDB website (http://cpdb.molgen.mpg.de/), resulting in identification of 383 genes and eight pathways. The Genotype–Tissue Expression database[15] was used to evaluate the impact of single-nucleotide polymorphisms (SNPs) on the regulation of target genes.

Exact Association Test (EXAT)

To identify genes associated with TB cavitation, we assumed that causal genes would exhibit distinct patterns of partial association between the case and control groups. In this study, we applied the EXAT, a statistical method designed to compare differences in variant association patterns between case and control groups using sequencing data.[16]

When applied to genome sequencing data for a specific gene, the data can be represented as contingency tables, reflecting variant distributions across study subjects. We assumed that disease-associated genes would exhibit differences in association between case and control groups. The EXAT test statistic was computed for each gene, and *P* values were obtained using a permutation-based procedure. Genes with a total read depth less than 10 or a minor allele frequency greater than 1%, non-coding genes, and pseudogenes excluded prior to analysis. Genes with *P* values less than the predefined significance threshold ($P < 0.05$) were identified as significantly associated with disease status.

Statistical analysis, data visualization, and network analysis

Fisher's exact test was performed to evaluate the enrichment of SNPs in the cavitation group. Genes with total read depth under 10 or minor allele frequency larger than 1%, non-coding genes and pseudogenes were eliminated before the analysis; all remaining genes were tested to confirm whether they follow Hardy–Weinberg equilibrium using the HardyWeinberg R package,[17] and SNPs with $P < 0.05$ were considered significantly deviated from equilibrium. All statistical analyses were conducted using R software (version 4.2; R Foundation for Statistical Computing, Vienna, Austria). Visualization of SNPs and related genes was carried out using the KaryoploteR[18] and Circlize R packages.[19] Network analysis was performed using Cytoscape version 3.9.1 (Institute of Systems Biology, Seattle, WA, USA) to explore the interactions between genes and pathways associated with cavitation.[20] To control for the false discovery rate, Benjamini–Hochberg–adjusted *P* values were calculated, with adjusted $P < 0.05$ considered statistically significant.

https://doi.org/10.3346/jkms.2026.41.e262

Ethics statement

The study protocol was approved by the Institutional Review Board of Hallym University Kangnam Sacred Heart Hospital (approval No. 2019-04-008). Written informed consent was obtained from all participants prior to enrollment.

RESULTS

Baseline characteristics

This study enrolled eleven TB patients with cavities and nine without cavities. The baseline characteristics of the study population are described in **Table 1**, comparing patients in the cavity group (n = 11) and the without-cavity group (n = 9). The median age was similar across the two groups (60.5, $P = 0.909$), and men were predominant (75.0%). Smoking status varied but was not significantly different ($P = 0.246$). Comorbidities included previous TB (15.0%) and diabetes mellitus (20.0%), with a greater prevalence of the latter in the cavity group (36.4% vs. 0%, $P = 0.144$). The median cavity size in the cavity group was 39.0 mm (interquartile range, 32.0–64.0 mm). Treatment outcomes showed a high cure rate (95.0%), with one case of follow-up loss (5.0%) and with no significant difference between groups ($P > 0.999$).

Genes associated with TB cavitation (EXAT results)

Table 2 summarizes the results of the EXAT analysis, identifying 30 genes associated with TB cavitation among the 383 genes analyzed. These genes were linked to key biological pathways involved in the development of cavitary TB, particularly those related to ECM remodeling and MMP activity. The most frequently represented pathway was ECM organization, encompassing 21 genes. This was followed by the ECM and the Rac1–Pak1–p38–MMP-2 signaling pathways, which included eight and seven genes, respectively. Notably, *MMP10* and *MMP3* were involved in four pathways and were most broadly associated with MMP-related signaling. *CTSK* and *KLK2* followed in prevalence, with each participating in three pathways. Genes such as *CAPN10* and *CAPNS2* were also implicated, pointing to additional pathways in ECM organization. Inflammatory and signaling mediators, including *TGFB1*, *MAPK3*, and *MAPK12*, were significantly enriched, indicating involvement of pro-inflammatory and apoptotic pathways in cavitary lesion development.[21-23]

Table 1. Baseline characteristics

Variables	Total (N = 20)	Cavity group (n = 11)	Without cavity group (n = 9)	*P* value
Age, yr	60.5 (43.0–68.5)	60.0 (46.5–68.0)	61.0 (41.0–68.0)	0.909
Sex, male	15 (75.0)	10 (90.9)	5 (55.6)	0.194
Smoking status				0.246
Never-smoker	9 (45.0)	4 (36.4)	5 (55.6)	
Ex-smoker	5 (25.0)	2 (18.2)	3 (33.3)	
Current smoker	6 (30.0)	5 (45.5)	1 (11.1)	
Comorbidities				
Previous tuberculosis	3 (15.0)	1 (9.1)	2 (22.2)	0.850
Diabetes mellitus	4 (20.0)	4 (36.4)	0 (0)	0.144
Chronic kidney disease	0 (0.0)	0 (0.0)	0 (0.0)	-
Cancer	0 (0.0)	0 (0.0)	0 (0.0)	-
Size of cavity, mm	24.0 (0.0–44.0)	39.0 (32.0–64.0)	-	-
Treatment outcome				1.000
Cured	19 (95.0)	10 (90.9)	9 (100.0)	
Follow-up loss	1 (5.0)	1 (9.1)	0 (0.0)	

Data are presented as number (%) or median (interquartile range).

https://doi.org/10.3346/jkms.2026.41.e262

Table 2. Genes associated with tuberculosis cavitation (Exact Association Test results)

Gene	*P* value	Adjusted *P* value[a]	Involved pathway(s)
MMP10	< 1.0e-5	< 1.0e-5	Matrix metalloproteinases Activation of matrix metalloproteinases Degradation of the extracellular matrix
MMP3	< 1.0e-5	< 1.0e-5	Extracellular matrix organization
CTSK	0.019	0.023	Activation of matrix metalloproteinases Degradation of the extracellular matrix
KLK2	< 1.0e-5	< 1.0e-5	Extracellular matrix organization
NCSTN	0.026	0.030	Degradation of the extracellular matrix
CAPN8	0.05	0.05	Extracellular matrix organization
CAPN10	< 1.0e-5	< 1.0e-5	
CAPNS2	< 1.0e-5	< 1.0e-5	
ACTN1	0.047	0.048	Extracellular matrix organization Cell–extracellular matrix interactions
TIMP3	< 1.0e-5	< 1.0e-5	Matrix metalloproteinases Inhibition of matrix metalloproteinases
MFAP2	< 1.0e-5	< 1.0e-5	Extracellular matrix organization
MATN3	< 1.0e-5	< 1.0e-5	
COL3A1	< 1.0e-5	< 1.0e-5	
KDR	0.009	0.012	
DDR1	< 1.0e-5	< 1.0e-5	
COL13A1	0.013	0.016	
EFEMP2	< 1.0e-5	< 1.0e-5	
ICAM2	0.002	0.003	
COMP	< 1.0e-5	< 1.0e-5	
TGFB1	< 1.0e-5	< 1.0e-5	
GDF5	< 1.0e-5	< 1.0e-5	
SDC4	< 1.0e-5	< 1.0e-5	
PIK3CA	0.031	0.035	Rac1–Pak1–p38–MMP-2 pathway
HRAS	0.031	0.034	
BAD	< 1.0e-5	< 1.0e-5	
MAPK3	< 1.0e-5	< 1.0e-5	
GRB7	< 1.0e-5	< 1.0e-5	
BAX	< 1.0e-5	< 1.0e-5	
MAPK12	< 1.0e-5	< 1.0e-5	
TNF	< 1.0e-5	< 1.0e-5	Matrix metalloproteinases

[a]Adjusted *P* value—false discovery rate correction.

SNPs enriched in patients with TB cavitation

A total of 35 SNPs associated with the presence of TB cavitation was identified using Fisher's exact test (**Table 3**). These SNPs are involved in various biological pathways, particularly those related to ECM organization and MMP activity. Of these, seven SNPs were associated significantly when a more stringent threshold of $P < 0.01$ was applied to reduce the risk of false discovery (**Fig. 1**). These significant SNPs were rs4253268 (regulating *KLKB1*), rs2076356 (*LAMB3*), rs58633563 and rs62107874 (*LTBP4*), rs140267288 (*LTBP4*), rs112115137 (*ITGB2*), and rs10241451 (*EGFR*) in lung tissue. Notably, rs4253268 (*KLKB1*), rs2076356 (*LAMB3*), and rs10241451 (*EGFR*) were associated with upregulation of their respective target genes. In contrast, the three SNPs regulating *LTBP4* (rs58633563, rs62107874, and rs140267288), along with rs112115137 (*ITGB2*), were predicted to result in the downregulation of gene expression. Functionally, *KLKB1* (rs4253268) is implicated in MMP activation and in pathways related to ECM degradation and organization. Similarly, *LAMB3* (rs2076356) is involved in ECM degradation and remodeling. *LTBP4* (rs140267288) and *ITGB2* (rs112115137) are associated with ECM organization. Finally, *EGFR* (rs10241451) is linked to the Rac1–Pak1–p38–MMP-2 signaling cascade.

https://doi.org/10.3346/jkms.2026.41.e262

Table 3. Single-nucleotide polymorphisms enriched in patients with tuberculosis cavitation (Fisher's exact test)

ID (dbSNP151)	Regulating target gene	*P* value (Fisher's exact test)	Variant effect for target gene expression from the GTEx database (slope)	Variant effect for target gene expression from GTEx database (*P* value)	Involved pathways
rs10088444	*MMP16*	0.026	0.326	5.51E-21	Matrix metalloproteinases Activation of matrix metalloproteinases
rs56187488	*MMP16*	0.026	0.307	3.87E-19	Degradation of the extracellular matrix
rs3741526	*MMP17*	0.038	−0.206	5.29E-9	Extracellular matrix organization
rs58370149	*KLKB1*	0.028	0.253	2.71E-5	Activation of matrix metalloproteinases Degradation of the extracellular matrix
rs4253268	*KLKB1*	0.010	0.244	3.80E-5	Extracellular matrix organization
rs2333618	*LAMC1*	0.028	0.145	7.91E-5	Degradation of the extracellular matrix
rs2076356	*LAMB3*	0.005	0.106	9.46E-7	Extracellular matrix organization
rs9429823	*LAMB3*	0.028	0.153	1.08E-12	
rs36080106	*CAPN8*	0.026	−0.104	4.06E-5	
rs4869306	*CAST*	0.038	0.169	2.30E-9	
rs760085	*COL9A3*	0.017	0.377	4.03E-9	
rs6586497	*MFAP2*	0.022	0.278	8.63E-18	Extracellular matrix organization
rs938511	*P3H2*	0.028	−0.175	1.00E-14	
rs72792241	*FBN2*	0.026	0.252	0.000160878	
rs13164401	*FBN2*	0.028	−0.225	0.000182489	
rs112573891	*SPARC*	0.028	−0.140	6.12E-5	
rs2881558	*SPARC*	0.038	−0.206	2.57E-12	
rs9396184	*COL21A1*	0.050	−0.170	0.000125388	
rs7860311	*MUSK*	0.038	0.157	1.87E-5	
rs61869681	*CD151*	0.026	−0.233	1.7805E-5	
rs497916	*TRAPPC4*	0.050	0.153	2.6081E-5	
rs58633563	*LTBP4*	0.005	−0.099	5.64E-5	
rs62107874	*LTBP4*	0.005	−0.095	0.00010014	
rs73047963	*LTBP4*	0.050	−0.101	1.24E-5	
rs140267288	*LTBP4*	0.008	−0.084	0.000107286	
rs2230528	*ITGB2*	0.050	0.102	1.98E-5	
rs112115137	*ITGB2*	0.005	−0.104	2.06E-7	
rs3859733	*ITGB2*	0.014	−0.096	7.74E-7	
rs10241451	*EGFR*	0.008	0.121	1.05E-6	Rac1–Pak1–p38–MMP-2 pathway
rs2740761	*EGFR*	0.050	0.130	1.88E-7	
rs73187236	*MAPK12*	0.026	−0.229	5.77E-5	
rs1009610	*MAPK12*	0.038	−0.266	4.06E-15	
rs112603021	*MAPK12*	0.038	−0.265	6.35E-15	
rs148762611	*MAPK12*	0.026	−0.187	1.43E-7	
rs73188924	*MAPK12*	0.026	−0.240	6.11E-9	

GTEx = Genotype–Tissue Expression.

An overview of MMP-related genes and SNPs associated with TB cavitation, visualized using KaryoploteR, is presented in **Fig. 2**, demonstrating that the genes identified by EXAT are not clustered on specific chromosomes but are distributed across the genome. A network analysis of cavitation-associated genes and MMP-related pathways, generated using Cytoscape, is shown in **Fig. 3**. Although the ECM organization and degradation pathways share a number of genes, the Rac1–Pak1–p38–MMP-2 signaling cascade appears to constitute a distinct and relatively independent cellular pathway.

DISCUSSION

In this study, we investigated host genetic factors involved in the development of cavitation in TB by performing small-scale WGS targeting MMP genes and adopting a gene-level association approach based on the EXAT methodology. Our analysis identified 30 genes and 35 SNPs that showed exploratory associations with cavitary lesions, many of which are

https://doi.org/10.3346/jkms.2026.41.e262

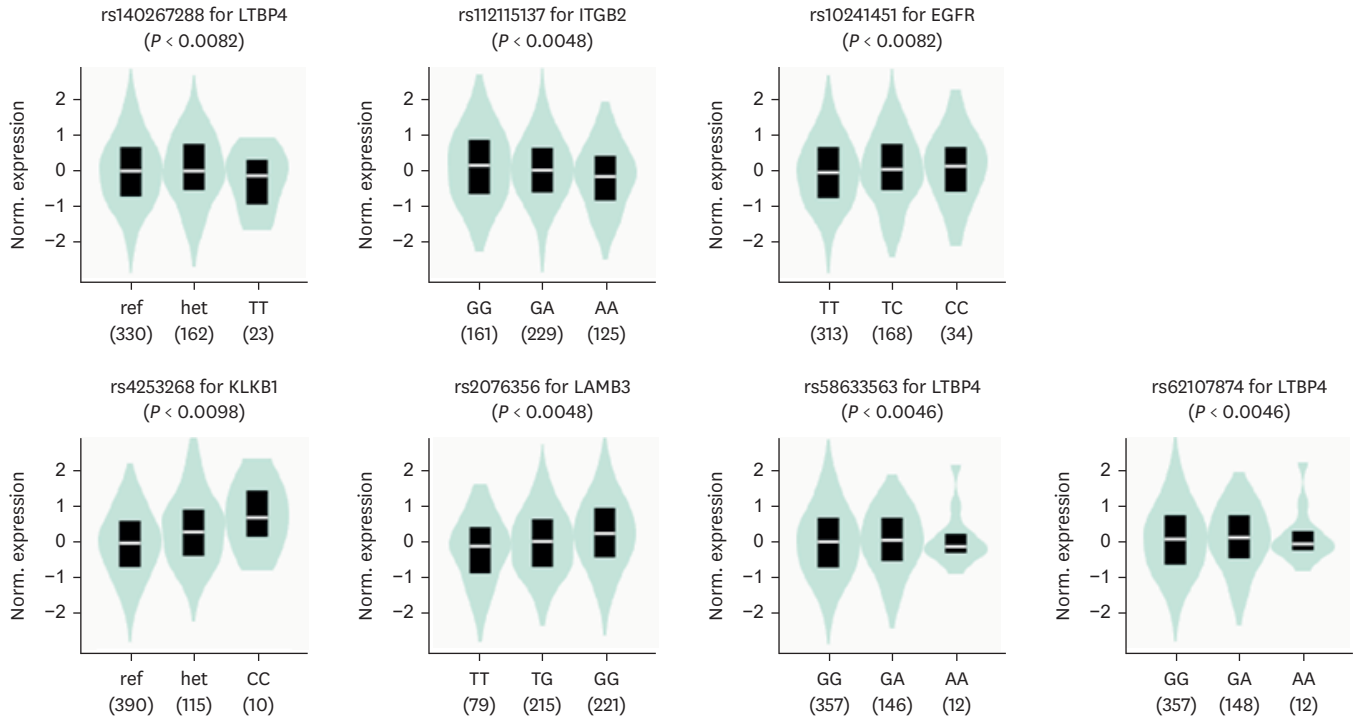

Fig. 1. Functional single-nucleotide polymorphisms with $P < 0.01$ in Fisher's exact test.

functionally implicated in the ECM organization, degradation, and activity of MMPs. To the best of our knowledge, this study provides an initial exploratory assessment of associations between TB cavitation and MMP-related genes and pathways.

In TB infection, granulomas often undergo liquefaction and necrosis, resulting in the formation of cavities, through enzymatic degradation of the ECM.[24-26] MMPs, particularly *MMP1*, *3*, *9*, *10*, and *11*, known to be secreted by MTB-infected monocytes, are central mediators of this process.[27] While most MMPs are minimally expressed under normal conditions, their expression is significantly upregulated during inflammation.[28] Experimental models have further demonstrated the role of *MMP3* in lung parenchymal destruction, reinforcing the connection between MMP activity and cavitary TB.[29,30]

Although the role of MMPs in the formation of cavitation in TB has been relatively well demonstrated, the genetic determinants influencing the expression and signaling of MMP-related genes remain poorly understood. From this view, our study provides informative data demonstrating that genes involved in various MMP-related pathways—particularly those linked to ECM degradation and organization—are significantly associated with TB cavitation. Notably, *MMP3* and *MMP10*, which participate in multiple MMP-related signaling cascades, exhibited notable exploratory differences between cavitary and non-cavitary TB patients. The simultaneous involvement of these genes in multiple MMP-related pathways further suggests a possible involvement in the pathogenesis of cavitary TB. In addition, SNPs in *KLKB1*, *LAMB3*, *LTBP4*, and *ITGB2,* the key regulators of MMP activity and ECM remodeling, also exhibited significant allelic frequency differences. These findings support the hypothesis that proteolytic breakdown of lung tissue, mediated by MMPs, may contribute to mechanisms of cavitation, consistent with prior evidence implicating MMPs in pulmonary tissue remodeling and destruction in TB.[6,30]

https://doi.org/10.3346/jkms.2026.41.e262

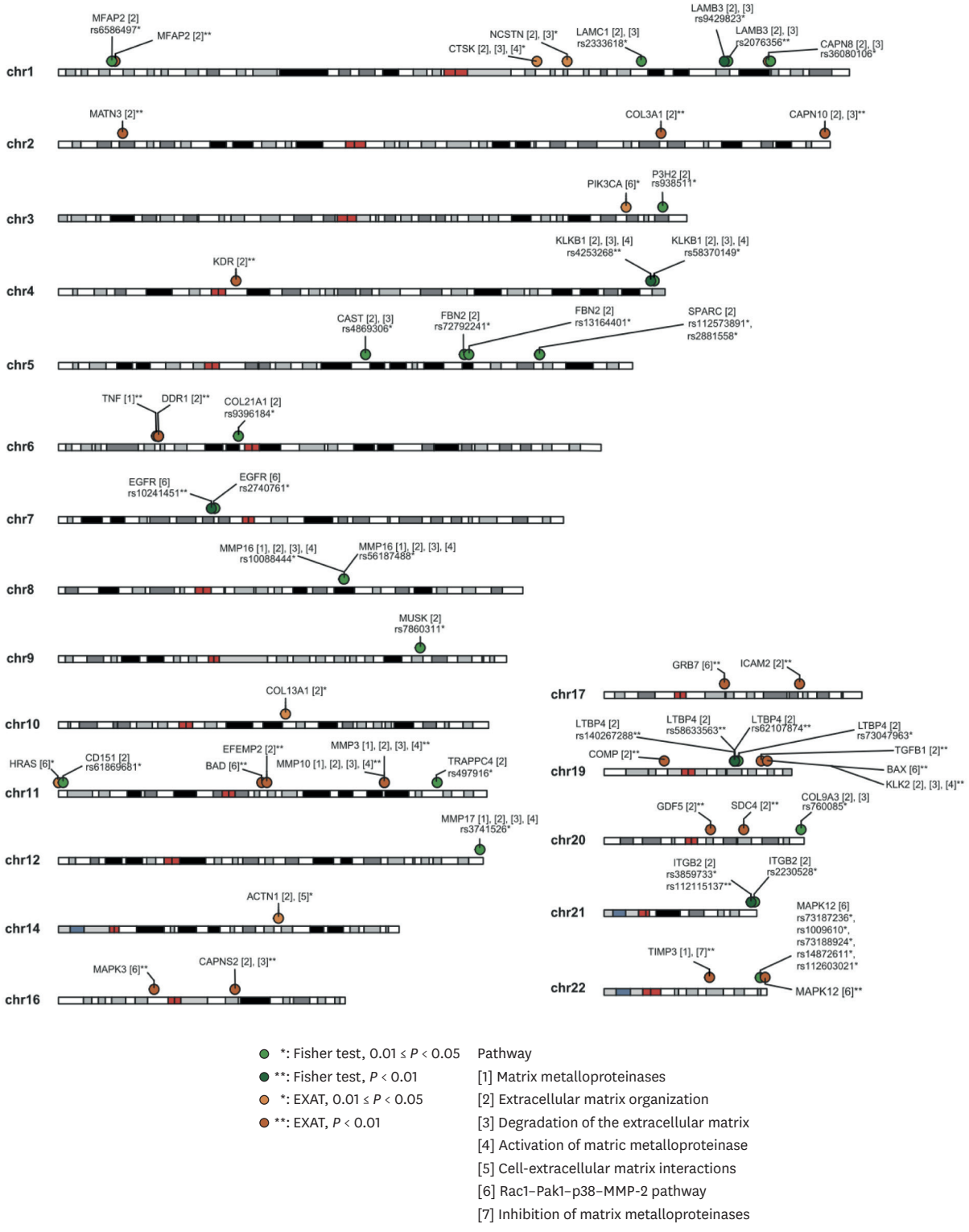

Fig. 2. Chromosomal distribution of cavitation-associated genes and SNPs across the human genome. The ideogram shows the genomic locations of 30 genes identified by EXAT analysis and 35 SNPs identified by Fisher's exact test that are significantly associated with tuberculosis cavity formation. Gene symbols and SNP identifiers (dbSNP151 rs numbers) are positioned according to their chromosomal coordinates based on the hg19 reference genome. Numbers in parentheses indicate the pathways in which each gene is involved (see legend). Statistical significance is denoted by asterisks ($^{*}P < 0.05$; $^{**}P < 0.01$) and color-coded circles: green circles represent results from Fisher's exact test and orange circles represent results from EXAT analysis. Darker shades indicate $P < 0.01$, whereas lighter shades indicate $0.01 \leq P < 0.05$.
SNP = single-nucleotide polymorphism, EXAT = Exact Association Test.

https://doi.org/10.3346/jkms.2026.41.e262

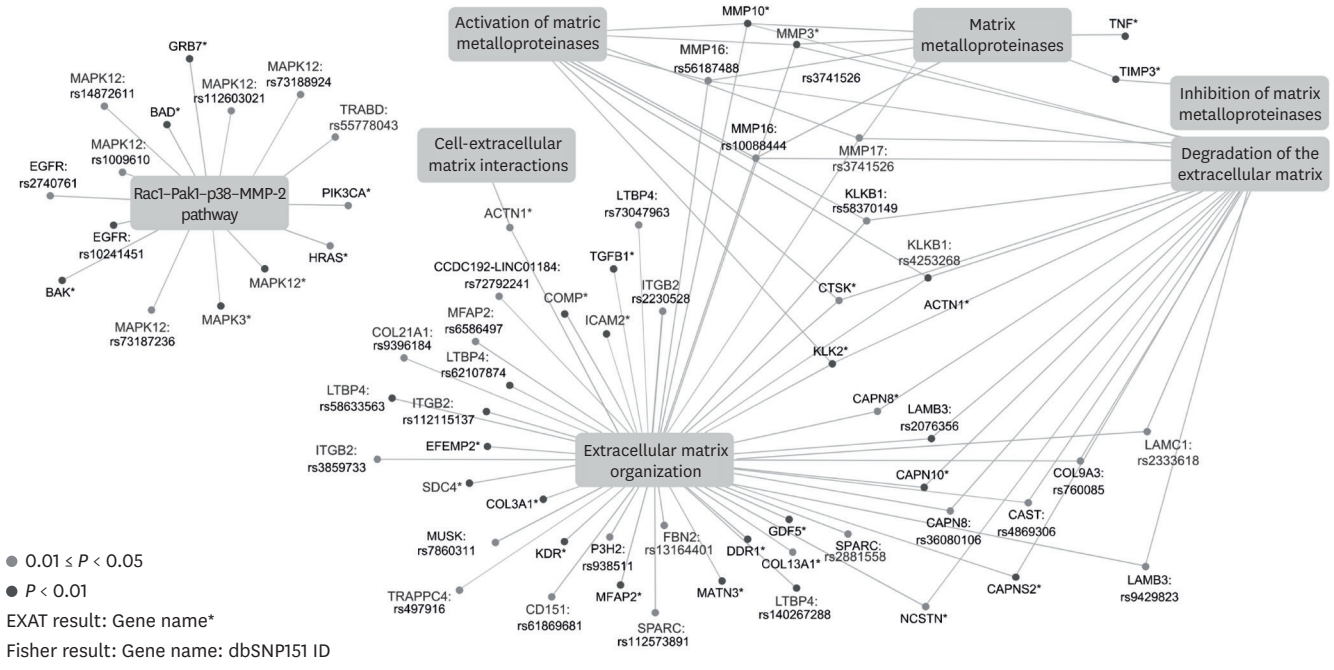

Fig. 3. Gene–pathway interaction network in cavitary tuberculosis. This network illustrates the functional relationships between host genetic factors and MMP-related pathways involved in tuberculosis cavity formation. The network was constructed using Cytoscape (v3.9.1; Institute of Systems Biology, Seattle, WA, USA) based on pathway annotation data from ConsensusPathDB. Large rounded rectangles represent biological pathways, while small circles represent genes (identified by Exact Association Test and marked with asterisks) and single-nucleotide polymorphisms (identified by Fisher's exact test and labeled with rs identifiers). Node shading indicates statistical significance (light: $0.01 \leq P < 0.05$; dark: $P < 0.01$). Edges denote gene–pathway associations. The Rac1–Pak1–p38–MMP-2 signaling pathway forms a distinct module (left), showing limited overlap with extracellular matrix degradation pathways, suggesting an independent regulatory mechanism. Hub genes (e.g., MMP3 and MMP10) participate in multiple pathways, indicating their pleiotropic roles in cavitary pathogenesis.
MMP = matrix metalloproteinase.

Furthermore, our study identifies an exploratory association between functional germline SNPs in *EGFR* and TB cavity formation (rs10241451, $P = 0.008$). *EGFR* activates Rac1-Pak1-p38 cascades implicated in MMP-2 regulation and ECM remodeling-pathways from cancer models,[31,32] that may inform TB tissue destruction, despite differing contexts. Convergent evidence includes germline EREG (an *EGFR* ligand) variants linked to TB susceptibility[33] and somatic *EGFR* mutations enriched in TB-associated adenocarcinoma (56% vs. 34%, $P = 0.038$).[34] We hypothesize that germline *EGFR* variants may modulate MMP activity, potentially influencing cavitation-requiring validation in TB models.

In addition, we observed enrichment of genes involved in pro-inflammatory and apoptotic signaling, including *TGFB1*, *TNF*, *MAPK3*, and *HRAS*, further supporting the notion that TB cavity formation is driven by a complex interplay among ECM degradation, inflammation, and cytoskeletal changes. These findings reinforce the idea of a multifactorial model in which immune-mediated tissue damage and proteolytic activity act synergistically. Moreover, karyogram analysis showed that associated genes were widely distributed across the genome rather than clustered on specific chromosomes, indicating the existence of a polygenic architecture without distinct chromosomal hotspots. The identification of multiple associated genes within individual patients suggests that susceptibility to cavitary TB is influenced by genetic heterogeneity, and that carrying a risk gene does not necessarily result in disease manifestation, possibly due to intrinsic genetic properties like incomplete penetrance or the multifactorial nature of the polygenic risk architecture. Acquired risk factors, including smoking and diabetes, may contribute to cavitary TB in conjunction with host genetic susceptibility.[35]

https://doi.org/10.3346/jkms.2026.41.e262

The clinical implications of these findings are important and multifaceted. The identification of specific genes and SNPs associated with TB cavitation provides insight into the molecular mechanisms underlying this severe form of TB. These genetic markers may offer preliminary clues regarding individual susceptibility to cavitary TB, enabling personalized treatment and management strategies. Additionally, these genetic insights generate hypotheses that may inform future research on potential therapeutic targets, that could improve outcomes for patients with cavitary TB.

Several limitations of this study should be acknowledged. First, the relatively small sample size may limit the generalizability of the findings and increase the risk of false positives despite multiple testing correction. In addition, we were unable to adequately adjust for potential confounders such as diabetes mellitus and smoking status due to the limited sample size, which may have influenced the associations observed. Second, functional validation of the identified genetic variants was not performed and should be addressed in future mechanistic studies. Third, our study focused exclusively on host genomic factors. Previous studies have reported that TB pathogenesis is influenced by interactions between host genetic susceptibility and *M. tuberculosis* strain diversity, affecting disease phenotype and severity.[36] Integration of host and bacterial genomic data may therefore offer a more comprehensive understanding of the host–pathogen interaction in cavitary TB. In addition, the lack of bacterial genomic data precluded lineage-specific analyses, which should be considered when interpreting the observed host genetic associations with cavitary disease. Furthermore, the observational nature of this study precludes causal inference between the identified genetic variants and cavitary disease development. Accordingly, the present findings should be interpreted as hypothesis-generating, and further validation through replication in larger, multi-center cohorts and integration with host–pathogen genomic and functional datasets will be required.

In conclusion, this study provides novel perspectives on host genetic factors associated with TB cavity formation. The identification of genes and SNPs linked to MMP activity, ECM remodeling, and immune signaling supports a multifactorial model of TB-associated tissue destruction. These findings offer preliminary, hypothesis-generating insights into host genetic contributors to cavitary TB. Further studies are needed to determine whether these signals have any relevance for biomarker development or host-directed therapeutic approaches.

REFERENCES

1. World Health Organization. Global Tuberculosis report 2024. Geneva, Switzerland: World Health Organization; 2024.
2. Choi H, Han K, Jung JH, Park SH, Kim SH, Kang HK, et al. Long-term mortality of tuberculosis survivors in Korea: a population-based longitudinal study. *Clin Infect Dis* 2023;76(3):e973-81. PUBMED | CROSSREF
3. Palaci M, Dietze R, Hadad DJ, Ribeiro FK, Peres RL, Vinhas SA, et al. Cavitary disease and quantitative sputum bacillary load in cases of pulmonary tuberculosis. *J Clin Microbiol* 2007;45(12):4064-6. PUBMED | CROSSREF
4. Benator D, Bhattacharya M, Bozeman L, Burman W, Cantazaro A, Chaisson R, et al. Rifapentine and isoniazid once a week versus rifampicin and isoniazid twice a week for treatment of drug-susceptible pulmonary tuberculosis in HIV-negative patients: a randomised clinical trial. *Lancet* 2002;360(9332):528-34. PUBMED | CROSSREF
5. Hamilton CD, Stout JE, Goodman PC, Mosher A, Menzies R, Schluger NW, et al. The value of end-of-treatment chest radiograph in predicting pulmonary tuberculosis relapse. *Int J Tuberc Lung Dis* 2008;12(9):1059-64. PUBMED

6. Brilha S, Chong DLW, Khawaja AA, Ong CWM, Guppy NJ, Porter JC, et al. Integrin α2β1 expression regulates matrix metalloproteinase-1-dependent bronchial epithelial repair in pulmonary tuberculosis. *Front Immunol* 2018;9:1348. PUBMED | CROSSREF
7. Kubler A, Larsson C, Luna B, Andrade BB, Amaral EP, Urbanowski M, et al. Cathepsin K contributes to cavitation and collagen turnover in pulmonary tuberculosis. *J Infect Dis* 2016;213(4):618-27. PUBMED | CROSSREF
8. Ong CW, Elkington PT, Friedland JS. Tuberculosis, pulmonary cavitation, and matrix metalloproteinases. *Am J Respir Crit Care Med* 2014;190(1):9-18. PUBMED | CROSSREF
9. Möller M, Hoal EG. Current findings, challenges and novel approaches in human genetic susceptibility to tuberculosis. *Tuberculosis (Edinb)* 2010;90(2):71-83. PUBMED | CROSSREF
10. Schurz H, Naranbhai V, Yates TA, Gilchrist JJ, Parks T, Dodd PJ, et al. Multi-ancestry meta-analysis of host genetic susceptibility to tuberculosis identifies shared genetic architecture. *eLife* 2024;13:e84394. PUBMED | CROSSREF
11. Raczy C, Petrovski R, Saunders CT, Chorny I, Kruglyak S, Margulies EH, et al. Isaac: ultra-fast whole-genome secondary analysis on Illumina sequencing platforms. *Bioinformatics* 2013;29(16):2041-3. PUBMED | CROSSREF
12. Cingolani P, Platts A, Wang LL, Coon M, Nguyen T, Wang L, et al. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly (Austin)* 2012;6(2):80-92. PUBMED | CROSSREF
13. Sherry ST, Ward MH, Kholodov M, Baker J, Phan L, Smigielski EM, et al. dbSNP: the NCBI database of genetic variation. *Nucleic Acids Res* 2001;29(1):308-11. PUBMED | CROSSREF
14. 1000 Genomes Project Consortium. A global reference for human genetic variation. *Nature* 2015;526(7571):68-74. PUBMED | CROSSREF
15. Lonsdale J, Thomas J, Salvatore M, Phillips R, Lo E, Shad S, et al. The Genotype-Tissue Expression (GTEx) project. *Nat Genet* 2013;45(6):580-5. PUBMED | CROSSREF
16. Lee J, Lee S, Jang JY, Park T. Exact Association Test for small size sequencing data. *BMC Med Genomics* 2018;11(S2 Suppl 2):30. PUBMED | CROSSREF
17. Graffelman J. Exploring diallelic genetic markers: the HardyWeinberg package. *J Stat Softw* 2015;64(3):1-23. CROSSREF
18. Gel B, Serra E. karyoploteR: an R/Bioconductor package to plot customizable genomes displaying arbitrary data. *Bioinformatics* 2017;33(19):3088-90. PUBMED | CROSSREF
19. Gu Z, Gu L, Eils R, Schlesner M, Brors B. *circlize* implements and enhances circular visualization in R. *Bioinformatics* 2014;30(19):2811-2. PUBMED | CROSSREF
20. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, et al. Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res* 2003;13(11):2498-504. PUBMED | CROSSREF
21. Deng Z, Fan T, Xiao C, Tian H, Zheng Y, Li C, et al. TGF-β signaling in health, disease, and therapeutics. *Signal Transduct Target Ther* 2024;9(1):61. PUBMED | CROSSREF
22. Morgan D, Berggren KL, Spiess CD, Smith HM, Tejwani A, Weir SJ, et al. Mitogen-activated protein kinase-activated protein kinase-2 (MK2) and its role in cell survival, inflammatory signaling, and migration in promoting cancer. *Mol Carcinog* 2022;61(2):173-99. PUBMED | CROSSREF
23. Ganguly P, Macleod T, Wong C, Harland M, McGonagle D. Revisiting p38 mitogen-activated protein kinases (MAPK) in inflammatory arthritis: a narrative of the emergence of MAPK-activated protein kinase inhibitors (MK2i). *Pharmaceuticals (Basel)* 2023;16(9):1286. PUBMED | CROSSREF
24. Parasa VR, Muvva JR, Rose JF, Braian C, Brighenti S, Lerm M. Inhibition of tissue matrix metalloproteinases interferes with *Mycobacterium tuberculosis*-induced granuloma formation and reduces bacterial load in a human lung tissue model. *Front Microbiol* 2017;8:2370. PUBMED | CROSSREF
25. Smigiel KS, Parks WC. Matrix metalloproteinases and leukocyte activation. *Prog Mol Biol Transl Sci* 2017;147:167-95. PUBMED | CROSSREF
26. Kaufmann SH, Dorhoi A. Inflammation in tuberculosis: interactions, imbalances and interventions. *Curr Opin Immunol* 2013;25(4):441-9. PUBMED | CROSSREF
27. Ong CW, Elkington PT, Brilha S, Ugarte-Gil C, Tome-Esteban MT, Tezera LB, et al. Neutrophil-derived MMP-8 drives AMPK-dependent matrix destruction in human pulmonary tuberculosis. *PLoS Pathog* 2015;11(5):e1004917. PUBMED | CROSSREF
28. Cui N, Hu M, Khalil RA. Biochemical and biological attributes of matrix metalloproteinases. *Prog Mol Biol Transl Sci* 2017;147:1-73. PUBMED | CROSSREF

29. Ravimohan S, Kornfeld H, Weissman D, Bisson GP. Tuberculosis and lung damage: from epidemiology to pathophysiology. *Eur Respir Rev* 2018;27(147):170077. PUBMED | CROSSREF

30. Xu Y, Wang L, Zimmerman MD, Chen KY, Huang L, Fu DJ, et al. Matrix metalloproteinase inhibitors enhance the efficacy of frontline drugs against *Mycobacterium tuberculosis*. *PLoS Pathog* 2018;14(4):e1006974. PUBMED | CROSSREF

31. Gonzalez-Villasana V, Fuentes-Mattei E, Ivan C, Dalton HJ, Rodriguez-Aguayo C, Fernandez-de Thomas RJ, et al. Rac1/Pak1/p38/MMP-2 axis regulates angiogenesis in ovarian cancer. *Clin Cancer Res* 2015;21(9):2127-37. PUBMED | CROSSREF

32. Binker MG, Binker-Cosen AA, Richards D, Oliver B, Cosen-Binker LI. EGF promotes invasion by PANC-1 cells through Rac1/ROS-dependent secretion and activation of MMP-2. *Biochem Biophys Res Commun* 2009;379(2):445-50. PUBMED | CROSSREF

33. Cao W, Luo LL, Chen WW, Liang L, Zhang RR, Zhao YL, et al. Polymorphism in the EREG gene confers susceptibility to tuberculosis. *BMC Med Genet* 2019;20(1):7. PUBMED | CROSSREF

34. Hwang IK, Paik SS, Lee SH. Impact of pulmonary tuberculosis on the EGFR mutational status and clinical outcome in patients with lung adenocarcinoma. *Cancer Res Treat* 2019;51(1):158-68. PUBMED | CROSSREF

35. Kim SH, Shin YM, Yoo JY, Cho JY, Kang H, Lee H, et al. Clinical factors associated with cavitary tuberculosis and its treatment outcomes. *J Pers Med* 2021;11(11):1081. PUBMED | CROSSREF

36. Aravindan PP. Host genetics and tuberculosis: theory of genetic polymorphism and tuberculosis. *Lung India* 2019;36(3):244-52. PUBMED | CROSSREF