

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our [Editorial Policies](#) and the [Editorial Policy Checklist](#).

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed
<input type="checkbox"/>	<input checked="" type="checkbox"/> The exact sample size ( $n$ ) for each experimental group/condition, given as a discrete number and unit of measurement
<input type="checkbox"/>	<input checked="" type="checkbox"/> A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
<input type="checkbox"/>	<input checked="" type="checkbox"/> The statistical test(s) used AND whether they are one- or two-sided <i>Only common tests should be described solely by name; describe more complex techniques in the Methods section.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of all covariates tested
<input type="checkbox"/>	<input checked="" type="checkbox"/> A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
<input type="checkbox"/>	<input checked="" type="checkbox"/> A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
<input type="checkbox"/>	<input checked="" type="checkbox"/> For null hypothesis testing, the test statistic (e.g. $F$ , $t$ , $r$ ) with confidence intervals, effect sizes, degrees of freedom and $P$ value noted <i>Give <math>P</math> values as exact values whenever suitable.</i>
<input type="checkbox"/>	<input checked="" type="checkbox"/> For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
<input checked="" type="checkbox"/>	<input type="checkbox"/> For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
<input type="checkbox"/>	<input checked="" type="checkbox"/> Estimates of effect sizes (e.g. Cohen's $d$ , Pearson's $r$ ), indicating how they were calculated

Our web collection on [statistics for biologists](#) contains articles on many of the points above.

Software and code

Policy information about [availability of computer code](#)

Data collection	Samples were genotyped using a customized Affymetrix LIMAArray (Luo et al, Nature Communications 2019) and GSA array. Mtb samples were sequenced on an Illumina HiSeq 2500 or 4000 sequencer and we called variants using Pilon (v1.22).
Data analysis	<p>All code for generating the figures presented in the manuscript are available at Luo, Y., Huang, C-C., Howard, N.C., Wang X., Liu QY et al. Paired analysis of host and pathogen genomes identifies determinants of human tuberculosis, <a href="https://github.com/yang-luo-lab/TB-g2g">https://github.com/yang-luo-lab/TB-g2g</a>, DOI: 10.5281/zenodo.13321932, 2024.</p> <p>We used custom scripts for preprocessing the human genotyping and Mtb sequencing data.</p> <p>We used the following command-line software and R package for statistical analysis: SAIGE (v1.1.6), <a href="https://saigegit.github.io/SAIGE-doc/">https://saigegit.github.io/SAIGE-doc/</a>; coloc (v3.2-1), <a href="https://cran.r-project.org/web/packages/coloc/index.html">https://cran.r-project.org/web/packages/coloc/index.html</a>; BEAST (v1.8.0), <a href="https://beast.community/beast">https://beast.community/beast</a>; Tracer (v1.7.0), <a href="https://beast.community/tracer">https://beast.community/tracer</a>; VEP (v.105), <a href="https://useast.ensembl.org/Tools/VEP">https://useast.ensembl.org/Tools/VEP</a>; PANTHER (v.16.0), <a href="http://pantherdb.org/">http://pantherdb.org/</a>; Pilon (v1.22), <a href="https://github.com/broadinstitute/pilon/wiki#documentation">https://github.com/broadinstitute/pilon/wiki#documentation</a>; IQ-TREE (v2), <a href="http://www.iqtree.org/">http://www.iqtree.org/</a>; Trimmomatic(v0.39), <a href="http://www.usadellab.org/cms/index.php?page=trimmomatic">http://www.usadellab.org/cms/index.php?page=trimmomatic</a>; STAR (v2.6.0), <a href="https://github.com/alexdobin/STAR/wiki">https://github.com/alexdobin/STAR/wiki</a>; RSEM (v1.2.29), <a href="https://deweylab.github.io/RSEM/">https://deweylab.github.io/RSEM/</a>;</p>

lme4 in R (v1.1-34), <https://www.rdocumentation.org/packages/lme4/versions/1.1-34/topics/lmer>;  
 xcms in R (v4.0.0), <https://bioconductor.org/packages/release/bioc/html/xcms.html>;  
 cluster in R (v2.1.4), <https://www.rdocumentation.org/packages/cluster/versions/2.1.4/topics/pam>

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio [guidelines for submitting code & software](#) for further information.

## Data

Policy information about [availability of data](#)

All manuscripts must include a [data availability statement](#). This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our [policy](#)

All code for generating the figures presented in the manuscript are available at Luo, Y., Huang, C-C., Howard, N.C., Wang X., Liu Q.Y., et al. Paired analysis of host and pathogen genomes identifies determinants of human tuberculosis, <https://github.com/yan-luo-lab/TB-g2g>, DOI: 10.5281/zenodo.13321932, 2024.

The human genotyping data generated in this study have been deposited in the dbGAP database under accession code phs002025.v1.p1 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs002025.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs002025.v1.p1)] and phs003718.v1 [[https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study\\_id=phs003718.v1.p1](https://www.ncbi.nlm.nih.gov/projects/gap/cgi-bin/study.cgi?study_id=phs003718.v1.p1)]

The Mycobacterium tuberculosis whole genome sequences in this study have been deposited in the BioProject database under accession code PRJNA1039243 [<https://www.ncbi.nlm.nih.gov/bioproject/?term=PRJNA1039243>].

The RNA sequencing data generated in this study have been deposited in the GEO database under accession code GSE262379 [<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE262379>].

The eQTL Catalogue release v6 database can be downloaded at <https://ftp.ebi.ac.uk/pub/databases/spot/eQTL/>.

The Genotype Tissue Expression (GTEx release v8) database can be downloaded at <https://gtexportal.org/home/protectedDataAccess>.

The whole-genome sequences of Mtb strains were obtained at the NCBI Sequence Read Archive (SRA) database (<https://www.ncbi.nlm.nih.gov/sra>) with Study ID listed in the Supplementary Data 10.

The Mtb reference assembly (H37Rv NC\_000962.3) can be downloaded at <https://www.ncbi.nlm.nih.gov/nuccore/CP003248>.

## Research involving human participants, their data, or biological material

Policy information about studies with [human participants or human data](#). See also policy information about [sex, gender \(identity/presentation\), and sexual orientation](#) and [race, ethnicity and racism](#).

Reporting on sex and gender	63% male and median age is 25 years. The sex of each participant was assigned based on genotyping and cross-checked with self-report data. No individual was excluded based on sex and age.
Reporting on race, ethnicity, or other socially relevant groupings	all individuals are from Lima, Peru, with admixed ancestry varying proportions among European-, African- and Native-American-genetic ancestry.
Population characteristics	age, gender and genetic principal component.
Recruitment	We enrolled index cases who presented with clinically suspected pulmonary TB at any of 106 participating health centers in Lima, Peru. Pulmonary TB patients have been diagnosed by the presence of acid fast bacilli in sputum smear or a positive Mtb culture at any time from enrollment to the end of treatment. This is a population cohort and we are
Ethics oversight	The study protocol was approved by the Institutional Review Board of Harvard School of Public Health and by the Research Ethics Committee of the National Institute of Health of Peru.

Note that full information on the approval of the study protocol must also be provided in the manuscript.

## Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences ☐ Behavioural & social sciences ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see [nature.com/documents/nr-reporting-summary-flat.pdf](https://www.nature.com/documents/nr-reporting-summary-flat.pdf)

## Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size We recruited 2,331 (1,632 discovery and 699 replication) individuals with tuberculosis disease from a large catchment area of Lima, Peru. We

Sample size	performed human host genotyping and bacterial whole genome sequencing to obtain paired genetic data used in this study. Assuming an additive effect using a chi-square test with a genome-wide significance threshold of $5 \times 10^{-8}$ for the power calculation, we estimated that we have >75% power to detect an interaction between a host allele of >10% frequency and a bacterial variant > 20% frequency with a relative genetic risk > 1.5.
Data exclusions	We excluded individuals that were missing more than 5% of the host genotype data, had an excess of heterozygous genotypes, and/or duplicated with identity-by-state > 0.9. We also excluded HIV positive individuals and bacterial isolates with evidence of mixed infection. Only individuals with paired genomic data are retained in the study.
Replication	All experimental measurement were taken the average of three replicates and multiple strains from the same testing lineages. All reported experimental findings are reproducible among biological replicates. The human replication cohort consists of 699 tuberculosis patients for reported interacting association, we reported a null statistical association in the replication cohort.
Randomization	The sampling was generally random from the described cohort where we enrolled patients who had a diagnosis of pulmonary TB at any of 106 participating health centers in Lima, Peru. We also recorded the index patients' baseline smear status, age, sex, HIV status, and drug-resistance profiles. We controlled for age, sex, genetic relationship and population structure in our association studies.
Blinding	No blinding was performed because we did not assign treatment groups.

## Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

### Materials & experimental systems

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> Antibodies
<input checked="" type="checkbox"/>	<input type="checkbox"/> Eukaryotic cell lines
<input checked="" type="checkbox"/>	<input type="checkbox"/> Palaeontology and archaeology
<input checked="" type="checkbox"/>	<input type="checkbox"/> Animals and other organisms
<input checked="" type="checkbox"/>	<input type="checkbox"/> Clinical data
<input checked="" type="checkbox"/>	<input type="checkbox"/> Dual use research of concern
<input checked="" type="checkbox"/>	<input type="checkbox"/> Plants

### Methods

n/a	Involved in the study
<input checked="" type="checkbox"/>	<input type="checkbox"/> ChIP-seq
<input checked="" type="checkbox"/>	<input type="checkbox"/> Flow cytometry
<input checked="" type="checkbox"/>	<input type="checkbox"/> MRI-based neuroimaging

## Plants

Seed stocks	<i>Report on the source of all seed stocks or other plant material used. If applicable, state the seed stock centre and catalogue number. If plant specimens were collected from the field, describe the collection location, date and sampling procedures.</i>
Novel plant genotypes	<i>Describe the methods by which all novel plant genotypes were produced. This includes those generated by transgenic approaches, gene editing, chemical/radiation-based mutagenesis and hybridization. For transgenic lines, describe the transformation method, the number of independent lines analyzed and the generation upon which experiments were performed. For gene-edited lines, describe the editor used, the endogenous sequence targeted for editing, the targeting guide RNA sequence (if applicable) and how the editor was applied.</i>
Authentication	<i>Describe any authentication procedures for each seed stock used or novel genotype generated. Describe any experiments used to assess the effect of a mutation and, where applicable, how potential secondary effects (e.g. second site T-DNA insertions, mosaicism, off-target gene editing) were examined.</i>