Genotypic and spatial analysis of transmission dynamics of tuberculosis in Shanghai, China: a 10-year prospective population-based surveillance study

Meng Li,[a,i] Liping Lu,[b,i] Qi Jiang,[a,c] Yuan Jiang,[d,e] Chongguang Yang,[a,f] Jing Li,[d,e] Yangyi Zhang,[d,e] Jinyan Zou,[b] Yong Li,[b] Wenqi Dai,[g] Jianjun Hong,[b] Howard Takiff,[h] Xin Shen,[d,e] Xiaoqin Guo,[b,j,***] Zhengan Yuan,[d,e,j,**] and Qian Gao[a,j,*]

[a]Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Science, Shanghai Medical College, Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, Shanghai, China
[b]Department of Tuberculosis Control, Songjiang District Center for Disease Control and Prevention, Shanghai, China
[c]School of Public Health, Renmin Hospital Public Health Research Institute, Wuhan University, Wuhan, China
[d]Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China
[e]Shanghai Institute of Preventive Medicine, Shanghai, China
[f]School of Public Health (Shenzhen), Shenzhen Campus of Sun Yat-sen University, Shenzhen, China
[g]Department of Clinical Laboratory, Songjiang District Central Hospital, Shanghai, China
[h]Laboratorio de Genética Molecular, CMBC, Instituto Venezolano de Investigaciones Científicas, IVIC, Caracas, Venezuela

The Lancet Regional Health - Western Pacific 2023;38: 100833

Published Online 29 June 2023
https://doi.org/10.1016/j.lanwpc.2023.100833

Summary

Background With improved tuberculosis (TB) control programs, the incidence of TB in China declined dramatically over the past few decades, but recently the rate of decrease has slowed, especially in large cities such as Shanghai. To help formulate strategies to further reduce TB incidence, we performed a 10-year study in Songjiang, a district of Shanghai, to delineate the characteristics, transmission patterns, and dynamic changes of the local TB burden.

Methods We conducted a population-based study of culture-positive pulmonary TB patients diagnosed in Songjiang during 2011–2020. Genomic clusters were defined with a threshold distance of 12-single-nucleotide-polymorphisms based on whole-genome sequencing, and risk factors for clustering were identified by logistic regression. Transmission inference was performed using phybreak. The distances between the residences of patients were compared to the genomic distances of their isolates. Spatial patient hotspots were defined with kernel density estimation.

Findings Of 2212 enrolled patients, 74.7% (1652/2212) were internal migrants. The clustering rate (25.2%, 558/2212) and spatial concentrations of clustered and unclustered patients were unchanged over the study period. Migrants had significantly higher TB rates but less clustering than residents. Clustering was highest in male migrants, younger patients and both residents and migrants employed in physical labor. Only 22.1% of transmission events occurred between residents and migrants, with residents more likely to transmit to migrants. The clustering risk decreased rapidly with increasing distances between patient residences, but more than half of clustered patient pairs lived ≥5 km apart. Epidemiologic links were identified for only 15.6% of clustered patients, mostly in close contacts.

Interpretation Although some of the TB in Songjiang's migrant population is caused by strains brought by infected migrants, local, recent transmission is an important driver of the TB burden. These results suggest that further reductions in TB incidence require novel strategies to detect TB early and interrupt urban transmission.

Funding Shanghai Municipal Science and Technology Major Project (ZD2021CY001), National Natural Science Foundation of China (82272376), National Research Council of Science and Technology Major Project of China (2017ZX10201302-006).

*Corresponding author. Key Laboratory of Medical Molecular Virology (MOE/NHC/CAMS), School of Basic Medical Science, Shanghai Medical College, Shanghai Institute of Infectious Disease and Biosecurity, Fudan University, 131 Dongan Road, 200032, Shanghai, China.
**Corresponding author. Shanghai Municipal Center for Disease Control and Prevention, Shanghai, China.
***Corresponding author. Department of Tuberculosis Control, Songjiang District Center for Disease Control and Prevention, Shanghai, China.
E-mail addresses: qiangao@fudan.edu.cn (Q. Gao), yuanzhengan@scdc.sh.cn (Z. Yuan), guoxiaoqin1102@163.com (X. Guo).
[i]These authors contributed equally to this work.
[j]These senior authors contributed equally to this work.

Keywords: Tuberculosis; Transmission patterns; Dynamic changes; Whole-genome sequencing; Spatial analysis

Research in context

Evidence before this study

We searched PubMed using the search terms "tuberculosis", "transmission", "migrant" and "China" from the inception of the database through March 31, 2023, for studies published in English. We found only 5 studies that used molecular epidemiology to investigate the origin and transmission characteristics of TB in large cities where the majority of TB patients were internal migrants. Searching the China national knowledge infrastructure (CNKI) with the same search terms for papers published in Chinese found no additional studies. China improved its TB control measures and reduced TB incidence over the past several decades, but despite good control programs, the rate of decline has slowed, especially in large cities. Long-term molecular epidemiological studies that delineate the dynamics of transmission should help formulate strategies to further reduce the incidence of TB in large cities, but we found no such studies.

Added value of this study

We combined genomic, epidemiological, and spatial analysis to investigate the patterns and dynamics of TB transmission in Songjiang, a district in Shanghai with a large population of migrant workers from rural China. The overall clustering rate of 25.2% was unchanged over the past decade and the percentage of clustered patients with confirmed epidemiological links was only 15.6%. However, while migrants had significantly higher TB rates and some migrants belonged to clusters, the clustering rate was considerably higher in long term residents of Songjiang. Only one fifth of transmission events occurred between residents and migrants, with residents more likely to transmit to migrants. The spatial concentration of TB patients was unchanged over the study period, and while patients living in close proximity had an increased risk of genomic clustering, more than half of clustered patients pairs lived ≥5 km apart.

Implications of all the available evidence

This study demonstrates that although much of the TB burden in Songjiang is caused by TB strains the migrants bring with them, this is superimposed on endemic recent, local transmission, mainly through casual contact, and these patterns have not changed over the past decade. These results suggest that the current TB prevention and control strategies, while good, may not be sufficient to achieve further reductions of the TB incidence in large cities, and therefore new strategies should be considered to enhance patient detection and interrupt transmission.

Introduction

Tuberculosis (TB) remains an important threat to global health and the second leading cause of death from an infectious disease after SARS-CoV-2.[1] China has the third highest TB burden worldwide and in 2021 had an estimated 780,000 new TB cases and 32,000 TB-related deaths.[1] China's rapid urbanization over the past several decades process has been accompanied by a vast migration of workers into large cities with employment opportunities.[2] As a result, areas that previously had low TB burdens must now cope with TB in migrants coming from TB high-burden areas.[3–5] Most internal migrants come from rural areas where the prevalence of TB is up to three times higher than in urban areas.[6] In addition, the internal migrants are more likely to have a low level of education, share crowded living conditions and have poor health awareness, all of which are risk factors for TB.[7] The internal migrants, who comprise more than 70% of new TB patients in large cities such as Shanghai and Shenzhen,[8,9] pose a daunting challenge to urban TB prevention and control. However, the TB brought by the migrants is superimposed upon the local TB burden that is mostly caused by transmission within the cities. Cluster analysis based on whole-genome sequencing (WGS) can help distinguish the proportion of the TB burden due to local transmission from the TB brought with the migrants, and thus inform targeted control strategies.[10,11]

We previously conducted genomic epidemiological studies of tuberculosis in Shanghai and Shenzhen to investigate the origin and transmission characteristics of TB in urban areas.[12,13] These prior studies, however, were of shorter duration or used lower resolution genotyping methods (VNTR). The more sensitive WGS produces a more accurate description of the local TB characteristics and transmission patterns, and a longer study duration makes it possible to determine whether these patterns have changed over time.

Shanghai is one of the most developed cities in China and since the 1990s has been a favored destination for rural-to-urban migration, especially into the Songjiang District, which was the first export industrial zone in Shanghai. In 2020 the population of Songjiang was 1.91 million, of whom 58.4% were internal migrants,[14] making it an appropriate, representative site for studying the epidemiology of TB in large Chinese cities. The incidence of TB in Shanghai has declined over the past few decades but more recently the rate of decrease has slowed, prompting an urgent need to find new strategies to achieve further reductions. Genomic epidemiological studies provide information on the nature of the local TB burden that can aid in the

formulation of effective, targeted control strategies. We therefore performed a study to delineate the patterns and dynamics of TB transmission in Songjiang, we prospectively collected TB isolates from Songjiang patients over a 10-year period for genomic, epidemiologic, and spatial analyses.

Methods

Study design and population

Community physicians routinely identify individuals with TB-like symptoms (cough for at least two weeks, fever, chest pain, weight loss, night sweats) or abnormal chest radiographs and refer them to the TB-designated hospital in Songjiang for diagnosis by sputum smear and culture. Sputum induction was used when patients could not produce sputum spontaneously. WGS was performed on all of the pre-treatment cultured isolates. The study population was comprised of all culture-positive pulmonary TB patients 15 years or older who were diagnosed between January 1, 2011 and December 31, 2020. The study was approved by the institutional review board of the Institutes of Biomedical Sciences, Fudan University and all enrolled patients provided informed written consent.

Whole-genome sequencing

All clinical strains were re-cultured and sequenced as described.[15] A previously validated pipeline was used to identify single nucleotide polymorphisms (SNPs).[16] In brief, raw sequence reads were trimmed with Sickle (version 1.33) and aligned to the inferred *Mycobacterium tuberculosis* complex ancestor sequence[17] using BWA-MEM.[18] SAMtools (version 1.3.1)[19] and Varscan (version 2.3.6)[20] were then used to identify SNPs. Strains with a sequence depth less than 20X or a genome coverage less than 95% were excluded from the analysis. Pairwise SNP distances were calculated based on the fixed SNPs with a frequency ≥75%. A genomic cluster was defined as strains differing by 12 or fewer SNPs, consistent with linkage through recent transmission.[21] Strain lineage identification and prediction of drug–resistance profiles were obtained from analyses using an online platform (https://samtb.uni-medica.com).[22]

Transmission inference

Transmission directions in each genomic cluster were inferred using the R package phybreak (version 0.5.2).[23] The priors used for mutation rate, the gamma-distributed generation and sampling time were as previously described.[24] We ran 20 independent Markov chain Monte Carlo (MCMC) simulations with a burn-in of 10,000 cycles and sampling of the independent chains every 50,000 cycles to ensure that most estimated parameters reached an effective sample size >200. Transmission directions with posterior probability >0.5 were included in further analysis.[24]

Spatial analysis

For all possible pairs of TB patients, we calculated the geographical distance between the residences given by the patients at the time of diagnosis using geosphere (version 1.5) in the R package. We estimated the SNP differences between all pairs of MTB isolates and evaluated the SNP differences at different geographic distances between the patient residences. Using more than 10 km as a reference, logistic regression was employed to calculate the odds ratios (OR) for genomic clustering at different levels of geographic proximity.[25] We also performed sensitivity analysis using 5 SNPs and 1 SNP as clustering thresholds. Kernel density estimation and spatial visualization were performed in ArcGIS (version 10.2).[26]

Epidemiological investigation

A questionnaire administered to all culture-positive TB patients within a week of their diagnosis asked about demographic, clinical and laboratory data. Information was also collected on the patients' close contacts, workplaces, residential addresses, and the social settings frequented in the three years prior to their TB diagnosis. Based on the WGS analysis, patients whose TB strains belonged to genomic clusters were invited to participate in an in-depth interview to identify epidemiological links with other patients in the same cluster. The epidemiological links were defined as confirmed when clustered patients knew each other and had a history of contact before TB diagnosis.

Statistical analysis

Non-normal continuous data was expressed as medians and interquartile ranges (IQR), while categorical variables were described using proportions. Differences between groups were tested using the Wilcoxon rank sum test or the chi-square test. Logistic regression was used to calculate the odds ratio (OR) and 95% confidence intervals (CI) for risk factors associated with genomic clustering. Variables with p-values less than 0.2 in the univariable analysis were included in the multivariable analysis to calculate the adjusted odds ratios (aOR). Factors with a p-value less than 0.05 in the final model were considered statistically significant. Changes in temporal trends of clustering rates were analyzed using Joinpoint (version 4.9.0.0).[27] All statistical analyses were performed in Stata version 14.0.

Role of the funding source

The funders played no role in the design or analysis of this study.

Results

Characteristics of study population

Between January 1, 2011 and December 31, 2020, there were 3012 bacteriologically confirmed pulmonary TB

patients in Songjiang, of which 2299 (76.3%) were culture-positive. Eighty-seven patients were excluded because their strains failed re-culture or genome sequencing. Of the remaining 2212 patients, 560 (25.3%) were residents and 1652 (74.7%) were internal migrants (Fig. 1A). Most migrant patients came to Songjiang from central and western regions of China (Fig. 1B), both of which have incidences of tuberculosis that are substantially higher than that of Shanghai.[28] Compared with resident patients, migrant patients were more likely to be female (33.7% vs 20.7%, $p < 0.001$), younger (27 years vs 58 years, $p < 0.001$), engaged in physical labor (77.6% vs 36.4%, $p < 0.001$), have multidrug-resistant isolates (5.4% vs 2.5%, $p = 0.008$), and were less likely to have a previous history of TB (4.7% vs 8.2%, $p = 0.002$) (Table 1).

Clustering rate and factors associated with genomic clustering

To estimate the level of recent transmission in Songjiang, we calculated the clustering rate between 2011 and 2020. A total of 558 (25.2%, 558/2212) strains were grouped into 200 genomic clusters containing 2–19 strains, and the cumulative clustering rate[16] gradually increased as more strains were analyzed (Fig. 2A/2B). To visually reflect temporal trends of the clustering rates in Songjiang, we calculated the clustering rates for sliding 3-year windows (Fig. 2C). Although the rates declined and then rose, there were no significant changes in the clustering rates over the course of the study period (Fig. 2D). We performed the same analysis separately for resident and migrant patients and again found no significant change in clustering rates during the study period (Supplementary Figure S1).

We then used logistic regression to identify risk factors associated with genomic clustering (Table 2). The multivariable analysis found that younger patients (<25 years: aOR 3.63, 95% CI 2.13–6.19; 25–44 years: aOR 2.86, 95% CI 1.71–4.78), males (aOR 1.59, 95% CI 1.26–1.99), and those engaged in physical labor (aOR 2.75, 95% CI 1.48–5.13) had a greater risk of clustering. However, compared with resident patients, migrant patients had an overall lower risk of clustering (aOR 0.42, 95% CI 0.31–0.56).

On further analysis we found that the risk factors for residents were somewhat different from those for migrants (Supplementary Tables S1 and S2). The clustering rates in the residents was highest, 40.6%, in the 30.4% of residents ≤44 years old and was 40.7% in the 36.4% of residents employed in physical labor. However, there was also a large portion (40.9%) of the resident TB patients who were ≥65 years old and had a much lower clustering rate of 18.3%. This suggests that while there is a younger, working resident population that acquired their TB from recent local transmission, there is also a large, older, resident population whose TB is unclustered and likely caused by reactivation of past infections.

Most of the TB in Songjiang (1395/2212, 63%) occurred in young internal migrants. In migrants, the highest levels of clustering were found in the <25 (28.1%) and 25–44 years old (24.2%) age groups, who made up 35.4% and 49.1% of the migrants, respectively. However, for all categories of migrant groups the clustering rates were lower than in the same categories of residents. This suggests that while there was clearly local transmission in the migrant population, it was less than in the residents, presumably because many of the migrants brought their TB infections with them from rural communities with a higher TB incidence.

Transmission between migrant and resident patients

To analyze the transmission between resident and migrant patients in Songjiang, we performed transmission inference for each genomic cluster. To include possible early source cases, the genomic clustering

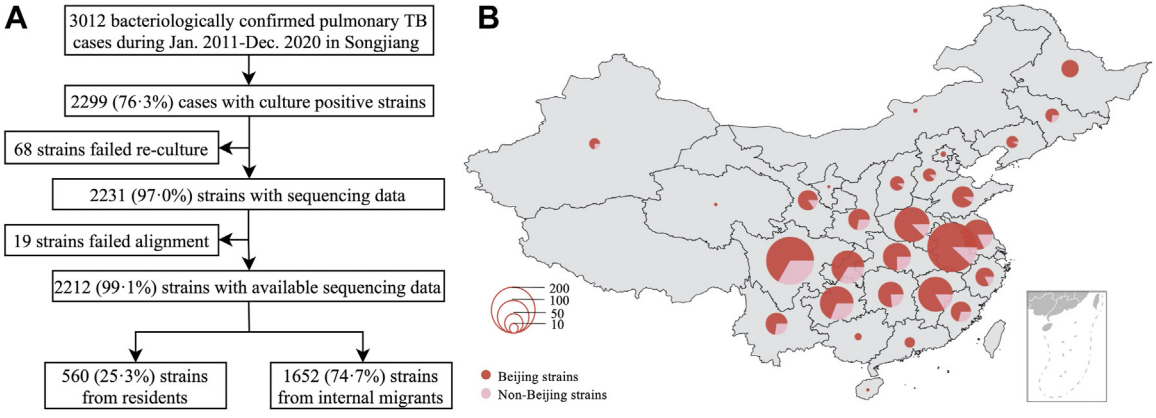

Fig. 1: Sample enrollment (A) and distribution of provinces of origin for migrant patients with tuberculosis in Songjiang, Shanghai, 2011–2020 (B). The size of the circles in (B) represents the number of patients, stratified by Beijing strains (red) and non-Beijing strains (pink).

analysis also included the WGS data of 61 cultured isolates obtained from Songjiang TB patients diagnosed in 2009 and 2010.[12] We found that 35 of these earlier patients were clustered with ten patients diagnosed in 2011–2020, for a total of 603 clustered patients grouped into 213 genomic clusters (Supplementary Table S3). After excluding the initial transmission event leading to the putative source case for each cluster, we estimated that there were 390 transmission events, of which 272 (69.7%) had a posterior probability >0.5. Of these 272 plausible transmission events, 47 occurred between resident patients and 165 between migrant patients. Only 60 (22.1%, 60/272) transmission events occurred between residents and migrants, with the residents were more likely to transmit to migrants than migrants transmitting to residents. Of the 196 transmission events initiated by a migrant, 31 (31/196, 15.8%) resulted in a secondary case in a resident, while out of the 76 transmission events initiated by a resident, 29 (29/76, 38.2%) resulted in a secondary case in a migrant ($p < 0.001$).

Epidemiological links of genomic-clustered patients

To delineate the transmission links between the genomic-clustered patients, we tried to perform in-depth epidemiological investigation of all clustered patients, and were able to complete the investigations for 67.2% (405/603) of these patients. The percentage of clustered patients with confirmed epidemiological links was 15.6% (94/603) overall, and was the same in both genomic-clustered residents (28/179, 15.6%) and migrants (66/424, 15.6%). The confirmed epidemiologic links were found to be predominantly with household contacts in both clustered resident (64.3%, 18/28) and migrant patients (53.0%, 35/66), and the rest were with friends, colleagues and neighbors. We identified only one confirmed link between a resident and a migrant patient who were workplace colleagues. Interestingly, two clusters included eight patients who worked in the same very large factory that employed many workers. However, they all labored in different workshops and did not know each other, so these were not considered to constitute confirmed epidemiological links.

Genetic relatedness and geographical distance

To look for a relationship between genomic clustering and spatial clustering of patient residences, we calculated the pairwise SNP differences and the geographical distances between the 2163 patients who provided detailed home addresses. We found a bimodal distribution of SNP differences, indicating closely related and more distantly related TB isolates, respectively, at all geographic distances between patient residences (Supplementary Figure S2). We then analyzed the geographic distance between the residences of genomic-clustered patient pairs and found that greater than half (55.8%, 419/751) of the patients in genomic-clusters lived more than 5 km apart (Fig. 3B), even when the clustering threshold distance was 5 SNPs or 1 SNP (Fig. 3C/3D). The median geographical distances between the residences of all paired genomic-clustered patients at clustering thresholds of 12 SNPs, 5 SNPs and 1 SNP, were 5.96 km, 5.94 km and 5.04 km (Supplementary Figure S3) respectively, which were not significantly different ($p = 0.1$).

We then analyzed the risk for genomic clustering at different levels of geographic proximity. We found that patients living within 1 km of each other had the highest risk of clustering (OR 14.68, 95% CI 11.54–18.66), but

	Migrant patients (n = 1652)	Resident patients (n = 560)	χ^2	p value
Sex			33.4024	<0.001
Female	557 (33.7)	116 (20.7)		
Male	1095 (66.3)	444 (79.3)		
Age (years), median	27 (23, 38)	58 (39, 72)	22.738	<0.001
Age group (years)				
<25	584 (35.4)	58 (10.4)	749.3409	<0.001
25–44	811 (49.1)	112 (20.0)		
45–64	219 (13.3)	161 (28.7)		
≥65	38 (2.3)	229 (40.9)		
Occupation			664.9174	<0.001
Retired	21 (1.3)	124 (22.1)		
Commercial service	93 (5.6)	12 (2.1)		
Physical labor	1282 (77.6)	204 (36.4)		
Farmer	39 (2.4)	138 (24.6)		
Students and teachers	50 (3.0)	34 (6.1)		
Other	167 (10.1)	48 (8.6)		
History of tuberculosis			9.6419	0.002
New cases	1574 (95.3)	514 (91.8)		
Retreated cases	78 (4.7)	46 (8.2)		
Diagnosis delay			9.0676	0.028
<2 weeks	382 (23.1)	98 (17.5)		
2–4 weeks	521 (31.5)	176 (31.4)		
4–8 weeks	455 (27.5)	172 (30.7)		
≥8 weeks	294 (17.8)	114 (20.4)		
Chest cavitation			0.1947	0.659
No	1134 (68.6)	390 (69.6)		
Yes	518 (31.4)	170 (30.4)		
Sputum smear status			0.1581	0.691
Negative	751 (45.5)	260 (46.4)		
Positive	901 (54.5)	300 (53.6)		
Drug resistance profile			9.5861	0.008
Drug-susceptible[a]	1442 (87.3)	494 (88.2)		
Non-MDR DR	121 (7.3)	52 (9.3)		
MDR[b]	89 (5.4)	14 (2.5)		
Beijing strain			29.7072	<0.001
No	344 (20.8)	59 (10.5)		
Yes	1308 (79.2)	501 (89.5)		

MDR: Multidrug-resistant. [a]Susceptible to the four first-line drugs. [b]Resistant to at least isoniazid and rifampicin.

Table 1: **Characteristics of migrant and resident TB patients in Songjiang, 2011–2020.**

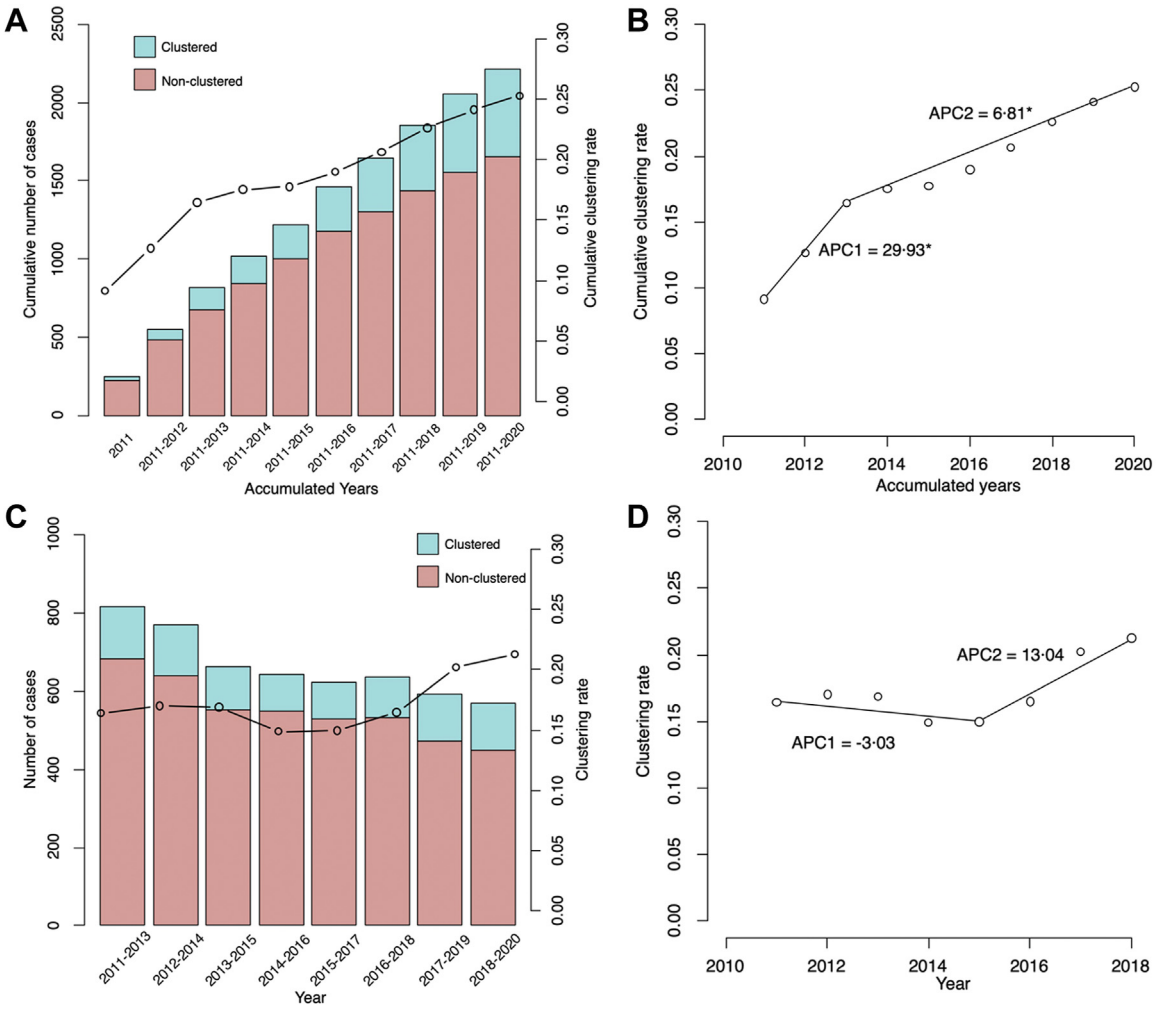

Fig. 2: Clustering rate and its changes in temporal trends. (A) Cumulative clustering rate in Songjiang. The bar indicates the cumulative number of cases, the line indicates the cumulative clustering rate. (B) Changes in temporal trends in cumulative clustering rate in Songjiang. *indicates that the Annual Percent Change (APC) is significantly different from zero at the alpha = 0.05 level. (C) Clustering rate of each sliding window 3 year period in Songjiang. The bar indicates the number of cases and the line indicates the clustering rate. (D) Changes in temporal trends in clustering rate for the sliding window 3 year period in Songjiang.

the risk decreased rapidly as the geographic distance between patient residences increased (Fig. 3A). These results remained consistent when the clustering threshold distance was reduced to 5 SNPs or 1 SNP (Fig. 3A, Supplementary Table S4).

Spatial hotspots and temporal changes

To identify spatial TB transmission hotspots, we used the residential address of the patients to draw kernel density maps. The spatial hotspots of clustered and unclustered patients were largely the same, concentrated in central and eastern Songjiang (Fig. 4C/4D). However, the resident and migrant patients showed distinct spatial distributions, with the resident patients concentrated in the central Songjiang (Fig. 4A) while the migrant patients were concentrated in areas of eastern Songjiang (Fig. 4B) – a distribution that reflects where most migrants and residents live in Songjiang.

Finally, we looked for temporal changes in the spatial hotspots by drawing kernel density maps of the residences of patients diagnosed in sliding 3-year windows, and found that the hotspots were largely unchanged over the 10-year duration of the study (Fig. 4E–L). We then stratified the analysis by clustered, unclustered, resident and migrant patients and again found no significant changes in the spatial distributions of the patient residences over the time-course of the study (Supplementary Figures S4–S7).

Discussion

This study, which was the longest longitudinal population-based genomic epidemiological study of TB

	Unclustered (n = 1654)	Clustered (n = 558)	Total (n = 2212)	Univariate regression		Multivariable regression	
				OR (95% CI)	p value	aOR (95% CI)	p value
Sex							
Female	535 (79.5)	138 (20.5)	673 (30.4)	1.00		1.00	
Male	1119 (72.7)	420 (27.3)	1539 (69.6)	1.46 (1.17, 1.81)	0.001	1.59 (1.26, 1.99)	<0.001
Age (years)							
<25	455 (70.9)	187 (29.1)	642 (29.0)	1.92 (1.34, 2.75)	<0.001	3.63 (2.13, 6.19)	<0.001
25–44	681 (73.8)	242 (26.2)	923 (41.7)	1.66 (1.18, 2.35)	0.004	2.86 (1.71, 4.78)	<0.001
45–64	298 (78.4)	82 (21.6)	380 (17.2)	1.29 (0.86, 1.92)	0.213	1.70 (1.05, 2.76)	0.031
≥65	220 (82.4)	47 (17.6)	267 (12.1)	1.00		1.00	
Occupation							
Retired	128 (88.3)	17 (11.7)	145 (6.6)	1.00		1.00	
Commercial service	80 (76.2)	25 (23.8)	105 (4.7)	2.35 (1.20, 4.63)	0.013	2.35 (1.09, 5.10)	0.03
Physical labor	1080 (72.7)	406 (27.3)	1486 (67.2)	2.83 (1.69, 4.75)	<0.001	2.75 (1.48, 5.13)	0.001
Farmer	132 (74.6)	45 (25.4)	177 (8.0)	2.57 (1.40, 4.72)	0.002	2.69 (1.45, 5.00)	0.002
Students and teachers	66 (78.6)	18 (21.4)	84 (3.8)	2.05 (0.99, 4.25)	0.052	1.29 (0.56, 2.96)	0.548
Other	168 (78.1)	47 (21.9)	215 (9.7)	2.11 (1.16, 3.84)	0.015	1.77 (0.89, 3.54)	0.106
History of tuberculosis							
New cases	1558 (74.6)	530 (25.4)	2088 (94.4)	1.00		–	
Retreated cases	96 (77.4)	28 (22.6)	124 (5.6)	0.86 (0.56, 1.32)	0.485	–	–
Internal migrant							
No	395 (70.5)	165 (29.5)	560 (25.3)	1.00		1.00	
Yes	1258 (76.2)	394 (23.8)	1652 (74.7)	0.76 (0.61, 0.94)	0.011	0.42 (0.31, 0.56)	<0.001
Diagnosis delay							
<2 weeks	363 (75.6)	117 (24.4)	480 (21.7)	1.00		–	
2–4 weeks	524 (75.2)	173 (24.8)	697 (31.5)	1.02 (0.78, 1.34)	0.862	–	–
4–8 weeks	456 (72.7)	171 (27.3)	627 (28.3)	1.16 (0.89, 1.53)	0.276	–	–
≥8 weeks	311 (76.2)	97 (23.8)	408 (18.4)	0.97 (0.71, 1.32)	0.835	–	–
Chest cavitation							
No	1124 (73.8)	400 (26.2)	1524 (68.9)	1.00		1.00	
Yes	530 (77.0)	158 (23.0)	688 (31.1)	0.84 (0.68, 1.03)	0.1	0.83 (0.66, 1.03)	0.09
Sputum smear status							
Negative	750 (74.2)	261 (25.8)	1011 (45.7)	1.00		–	
Positive	904 (75.3)	297 (24.7)	1201 (54.3)	0.94 (0.78, 1.14)	0.558	–	–
Drug resistance profile							
Pan-susceptible	1427 (73.7)	509 (26.3)	1936 (87.5)	1.00		1.00	
Non-MDR DR	144 (83.2)	29 (16.8)	173 (7.8)	0.56 (0.37, 0.85)	0.006	0.57 (0.37, 0.87)	0.009
MDR	83 (80.6)	20 (19.4)	103 (4.7)	0.68 (0.41, 1.11)	0.123	0.61 (0.37, 1.02)	0.06
Beijing strain							
No	346 (85.9)	57 (14.1)	403 (18.2)	1.00		1.00	
Yes	1308 (72.3)	501 (27.7)	1809 (81.8)	2.33 (1.73, 3.13)	<0.001	2.28 (1.68, 3.09)	<0.001

Table 2: **Univariate and multivariable logistic regression of risk factors for genomic clustering.**

in urban areas in China, combined genomic, epidemiological, and spatial analysis to investigate the patterns and dynamics of TB transmission in Songjiang. Over the 10 year duration of the study, we found no significant changes in either the clustering rate or the spatial distribution of the residences of TB patients in Songjiang. While patients living in close proximity had an increased risk of genomic clustering, this risk decreased rapidly with increasing distance between patient residences.

It was thought that WGS would identify transmission hotspots as targets for strategic interventions. Consistent with previous findings,[12,15,25] the closer the geographical distance between the residences of Songjiang patients, the higher the risk of clustering, but the geographic locations with the most clustering were simply the regions of Songjiang where most migrants, including most unclustered migrant TB patients, resided, and therefore no real transmission hotspots could be identified. Epidemiologic links could be found for only 15.6% of the clustered patients, much lower than the 41.8% in rural areas,[16] and most of the links were with household contacts. In addition, over half of

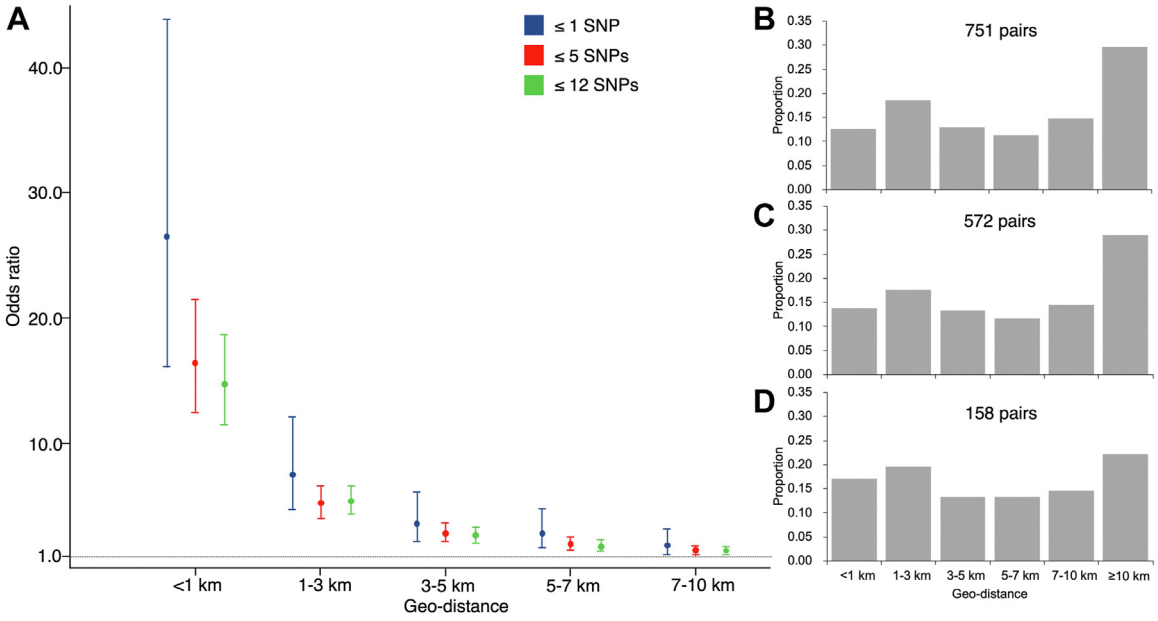

Fig. 3: Associations between pairwise geographic distance and genomic clustering using pairs located 10 or more kilometers apart as the reference category (A). Different colors indicate different thresholds for genomic clustering. Error bars are 95% confidence intervals. Histograms by geographic distance for pairs with (B) ≤12 SNPs difference, (C) ≤5 SNPs difference, and (D) ≤1 SNP difference.

the clustered patient pairs lived more than 5 km from each other and had no known connections, suggesting that most clustered cases were infected by unknown, casual contacts. Of the 272 plausible transmission events, most occurred within each group and only 60 (22%) were between residents and migrants, with a higher proportion of transmission from residents to migrants than the reverse. In areas where TB transmission primarily occurs through casual contacts, establishing a transmission monitoring system is crucial. This system facilitates the prompt detection and investigation of outbreaks, enabling early intervention to interrupt transmission. Furthermore, collecting comprehensive social network information from patients is essential for assessing the association between exposure to community venues and TB. Intervention across the segregated network of case venues may be necessary to effectively interrupt transmission.[29]

The overall clustering rate in Songjiang was 25.2%, significantly higher than in Shenzhen, China (12.2%) or Oxfordshire, UK (15.8%), where the majority of TB patients were also migrants from high-burden areas.[3,13] The 10-year duration of the study was long enough to have captured most of the TB from recent transmission in Songjiang during this period, although the clustering rate in the residents would likely have been higher if it included the TB cases diagnosed in all of Shanghai during the study period. In addition, some of the cases of TB in young residents were probably reactivation of distantly acquired infections that, as in the older residents, were less likely to be clustered. The study's temporal and spatial limitations could have also affected the clustering rate in migrants, who were mostly young and employed in physical labor, but the underestimation likely had a smaller impact due to the shorter migrant residency in Shanghai. The highest clustering rate in migrants, 28.1% in those <25 years (overall 23.8%), was well below the highest resident rate in physical laborers (40.7%), and lower than the overall clustering rate for all residents (29.3%). This suggests that while there is substantial local transmission among migrants in Songjiang, a large portion of migrant TB was probably due to reactivation of infections with TB strains acquired in their hometowns.

The highest levels of clustering, in both residents and migrants, were in the physical laborers who would be expected to have lower socio-economic status and were likely infected either at their worksites or as they moved around Shanghai. The clustering of eight patients, who worked in different areas of the same large factory and didn't know each other, could be considered a microcosm of transmission in Shanghai. Although it seems perhaps surprising that patients with active TB could perform physical labor, individuals in this younger, more healthy group[7] might resist early symptoms or transmit their TB during an infectious preclinical stage of the disease.[30]

Shanghai is one of the most developed cities in China and has one of the best TB prevention and control programs. It was the first city in China to implement a series of pioneering strategies for TB diagnosis, treatment and control, and saw a decline in the notified

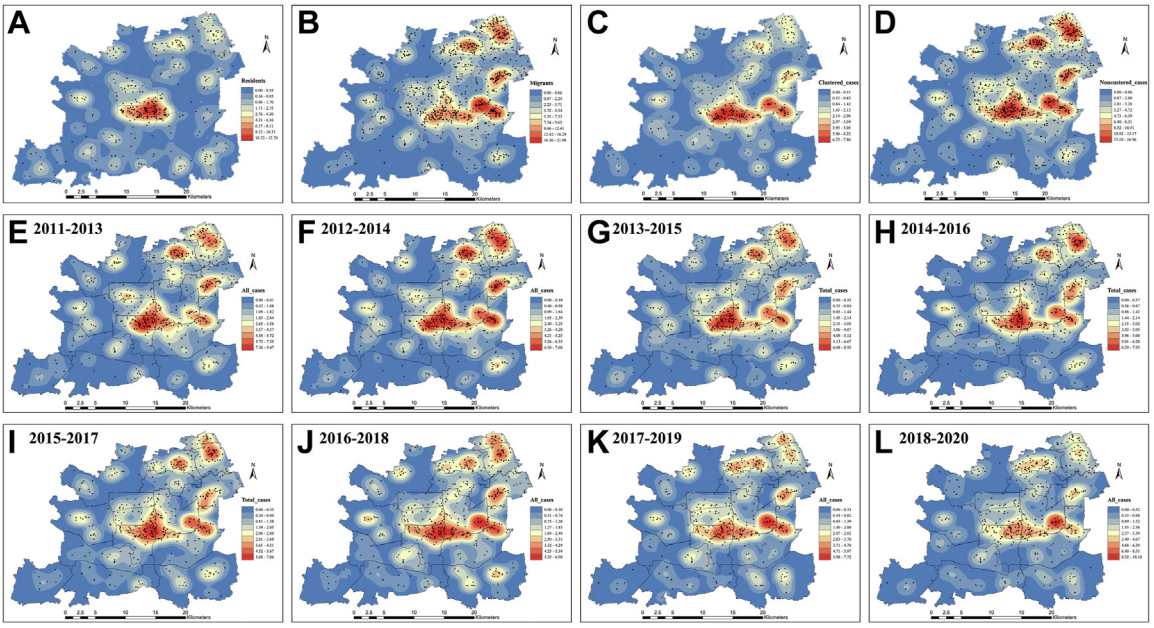

Fig. 4: Kernel density maps of resident patients (A), migrant patients (B), clustered patients (C), unclustered patients (D), and all patients in 3 year sliding window intervals (E–L).

incidence from 64 per 100,000 population in the late 1990s to less than 30 per 100,000 population today.[31–34] However, our previous study showed that this decline slowed over the past decade (Supplementary Figure S8), with the notified incidence remaining around 25 per 100,000 in the resident population and around 32 per 100,000 population in migrants.[35] In the current study, we found no significant changes in the clustering rate or spatial distribution of TB patients in Songjiang from 2011 to 2020, suggesting there were no significant reductions in recent, local transmission. Compared to countries with low TB burdens, such as the United States (2.6 per 100,000 population), the United Kingdom (6.3 per 100,000 population), or Japan (11 per 100,000 population),[1] the TB burden of Shanghai remains high. However, while our study could not detail the transmission chains for most patients, nor identify transmission hotspots, it nevertheless provided valuable insights into the local epidemiology of the disease. The clustering analysis revealed that recent transmission remains an important driver of tuberculosis infection in Songjiang and has not been significantly reduced even by the implementation of good, standard TB control measures. Achieving further significant reductions in the TB incidence in Shanghai and other Chinese cities may require additional targeted strategies based on the characteristics of the local TB burden and the high risk populations.[36]

The most important limitation of this study is that the calculated clustering rate likely underestimates the extent of local TB transmission because strains could be misclassified as unique when they were, in fact, clustered with strains outside the study's temporal or geographical limits. A previous study of MDR-TB patients across the entire city of Shanghai found a clustering rate of 32%,[21] but when only patients from Songjiang were included the clustering rate was just 19% ($p = 0.02$), indicating substantial cross-district transmission. In addition, our geographical data included only the locations where patients resided and did not analyze their patterns of mobility within the city.

In conclusion, neither the clustering rate nor the spatial distribution of clustering in Songjiang TB patients has changed significantly over the past decade. While many migrants bring their TB strains with them, either as latent or subclinical infections, local transmission is an important driver of TB in both the residents and migrants in Songjiang. The risk for genomic clustering decreased rapidly with increasing geographical distances between the residences, and most clustered cases lived far apart and had no identifiable epidemiological links. The current TB prevention and control strategies, while good, may not be sufficient to achieve further reductions in the TB incidence in Shanghai. Therefore new strategies, based on the specific characteristic of the local TB burden, should be considered in order to enhance case detection and thereby interrupt transmission.

Contributors

QJ, CY, XS, XG, ZY and QG designed and managed the study. ML, LL, JZ, YL, WD and JH performed the epidemiological investigation; YJ, JL, YZ and YL performed the experiments and collected the laboratory data;

ML and QJ cleaned the data and performed statistical analysis; ML performed the sequence analysis and interpretation. ML, QJ, CY, HT and QG prepared the manuscript. All authors contributed to and gave input to the final version of the manuscript.

Data sharing statement

Sequencing data were deposited in the Genome Sequence Archive (https://bigd.big.ac.cn/gsa) under BioProject PRJCA010372. De-identified participant data from the study will be made available upon publication to medical researchers on a not-for-profit basis by email request to the corresponding author for the purposes of propensity matching or meta-analysis.

Editor note

The Lancet Group takes a neutral position with respect to territorial claims in published maps and institutional affiliations.

Declaration of interests

The authors have declared that no competing interests exist.

Acknowledgements

We thank the tuberculosis public health teams in the Songjiang District Center for Disease Control and Prevention.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.lanwpc.2023.100833.

References

1 WHO. *Global tuberculosis report 2022*. Geneva: World Health Organization; 2022.
2 National Bureau of Statistics. *Survey on migrant workers 2020*. Beijing, China: National Bureau of Statistics; 2020.
3 Walker TM, Lalor MK, Broda A, et al. Assessment of Mycobacterium tuberculosis transmission in Oxfordshire, UK, 2007-12, with whole pathogen genome sequences: an observational study. *Lancet Respir Med*. 2014;2(4):285–292.
4 Lonnroth K, Mor Z, Erkens C, et al. Tuberculosis in migrants in low-incidence countries: epidemiology and intervention entry points. *Int J Tuberc Lung Dis*. 2017;21(6):624–637.
5 Tagliani E, Anthony R, Kohl TA, et al. Use of a whole genome sequencing-based approach for Mycobacterium tuberculosis surveillance in Europe in 2017-2019: an ECDC pilot study. *Eur Respir J*. 2021;57(1):2002272.
6 Wang L, Zhang H, Ruan Y, et al. Tuberculosis prevalence in China, 1990-2010;a longitudinal analysis of national survey data. *Lancet*. 2014;383(9934):2057–2064.
7 Hu X, Cook S, Salazar MA. Internal migration and health in China. *Lancet*. 2008;372(9651):1717–1719.
8 Shen X, Xia Z, Li X, et al. Tuberculosis in an urban area in China: differences between urban migrants and local residents. *PLoS One*. 2012;7(11):e51133.
9 Wu Q, Lyu D, Guan H, et al. Analysis of epidemiological characteristics of pulmonary tuberculosis in Shenzhen from 2007 to 2016. *J Trop Med*. 2018;18(1):86–89.
10 Alland D, Kalkut GE, Moss AR, et al. Transmission of tuberculosis in New York City. An analysis by DNA fingerprinting and conventional epidemiologic methods. *N Engl J Med*. 1994;330(24):1710–1716.
11 Small PM, Hopewell PC, Singh SP, et al. The epidemiology of tuberculosis in San Francisco. A population-based study using conventional and molecular methods. *N Engl J Med*. 1994;330(24):1703–1709.
12 Yang C, Lu L, Warren JL, et al. Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis. *Lancet Infect Dis*. 2018;18(7):788–795.
13 Yang T, Wang Y, Liu Q, et al. A population-based genomic epidemiological study of the source of tuberculosis infections in an emerging city: Shenzhen, China. *Lancet Reg Health-W*. 2021;8:100106.
14 Shanghai Bureau of Statistics. *ShangHai statistical yearbook 2021*. Shanghai Bureau of Statistics; 2021.
15 Jiang Q, Liu Q, Ji L, et al. Citywide transmission of multidrug-resistant tuberculosis under China's rapid urbanization: a retrospective population-based genomic spatial epidemiological study. *Clin Infect Dis*. 2020;71(1):142–151.
16 Li M, Guo M, Peng Y, et al. High proportion of tuberculosis transmission among social contacts in rural China: a 12-year prospective population-based genomic epidemiological study. *Emerg Microbes Infect*. 2022;11(1):2102–2111.
17 Comas I, Coscolla M, Luo T, et al. Out-of-Africa migration and Neolithic coexpansion of Mycobacterium tuberculosis with modern humans. *Nat Genet*. 2013;45(10):1176–1182.
18 Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics*. 2009;25(14):1754–1760.
19 Li H. A statistical framework for SNP calling, mutation discovery, association mapping and population genetical parameter estimation from sequencing data. *Bioinformatics*. 2011;27(21):2987–2993.
20 Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568–576.
21 Yang C, Luo T, Shen X, et al. Transmission of multidrug-resistant Mycobacterium tuberculosis in Shanghai, China: a retrospective observational study using whole-genome sequencing and epidemiological investigation. *Lancet Infect Dis*. 2017;17(3):275–284.
22 Yang T, Gan M, Liu Q, et al. SAM-TB: a whole genome sequencing data analysis website for detection of Mycobacterium tuberculosis drug resistance and transmission. *Brief Bioinform*. 2022;23(2):bbac030.
23 Klinkenberg D, Backer JA, Didelot X, Colijn C, Wallinga J. Simultaneous inference of phylogenetic and transmission trees in infectious disease outbreaks. *PLoS Comput Biol*. 2017;13(5):e1005495.
24 Gygli SM, Loiseau C, Jugheli L, et al. Prisons as ecological drivers of fitness-compensated multidrug-resistant Mycobacterium tuberculosis. *Nat Med*. 2021;27(7):1171–1177.
25 Huang CC, Trevisi L, Becerra MC, et al. Spatial scale of tuberculosis transmission in Lima, Peru. *Proc Natl Acad Sci U S A*. 2022;119(45):e2207022119.
26 Zelner JL, Murray MB, Becerra MC, et al. Identifying hotspots of multidrug-resistant tuberculosis transmission using spatial and molecular genetic data. *J Infect Dis*. 2016;213(2):287–294.
27 Kim HJ, Fay MP, Feuer EJ, Midthune DN. Permutation tests for joinpoint regression with applications to cancer rates. *Stat Med*. 2000;19(3):335–351.
28 Jiang H, Liu M, Zhang Y, et al. Changes in incidence and epidemiological characteristics of pulmonary tuberculosis in mainland China, 2005-2016. *JAMA Netw Open*. 2021;4(4):e215302.
29 Bui DP, Oren E, Roe DJ, et al. A case-control study to identify community venues associated with genetically-clustered, multidrug-resistant tuberculosis disease in Lima, Peru. *Clin Infect Dis*. 2019;68(9):1547–1555.
30 Ryckman TS, Dowdy DW, Kendall EA. Infectious and clinical tuberculosis trajectories: bayesian modeling with case finding implications. *Proc Natl Acad Sci U S A*. 2022;119(52):e2211045119.
31 Zhang S, Yuan Z, Mei J, Shen M, Shen X. The effectiveness of the new control network in Shanghai. *Chin J Antituberc*. 2007;(1):74–77.
32 Mei J. Tuberculosis prevention and treatment is facing a new historical turning point in Shanghai. *Shanghai J Prevent Med*. 2019;31(1):23–27.
33 Shen X, Pan Q, Mei J, Yuan Z, Wu G, Chen X. The improvement of free treatment policy for pulmonary tuberculosis in Shanghai. *Chin Health Res*. 2019;22(4):266–268.
34 Wang H, Shen X, Chen J, Xia Z, Xu B, Yuan Z. Analysis of the epidemiologieal characteristics of pulmonary tuberculosis among migrants in Shanghai from 2008 to 20l9. *Chin J Antituberc*. 2021;43(4):370–377.
35 Zou J, Lu L, Li Y, Jin L. Analysis of the epidemiological characteristics of pulmonary tuberculosis among nonlocal people in Songjiang District of Shanghai City from 2010 to 2019. *J Tuberc Lung Dis*. 2022;3(3):236–241.
36 Lu L, Li M, Chen C, et al. Outbreak of tuberculosis in internet cafes amongst young internal migrants without fixed abode in Shanghai, China, 2018-2019. *J Travel Med*. 2023;30(1):taac121.