Internal migration and transmission dynamics of tuberculosis in Shanghai, China: an epidemiological, spatial, genomic analysis

Chongguang Yang, Liping Lu, Joshua L Warren, Jie Wu, Qi Jiang, Tianyu Zuo, Mingyu Gan, Mei Liu, Qingyun Liu, Kathryn DeRiemer, Jianjun Hong, Xin Shen, Caroline Colijn, Xiaoqin Guo*, Qian Gao*, Ted Cohen*

Summary

Background Massive internal migration from rural to urban areas poses new challenges for tuberculosis control in China. We aimed to combine genomic, spatial, and epidemiological data to describe the dynamics of tuberculosis in an urban setting with large numbers of migrants.

Methods We did a population-based study of culture-positive *Mycobacterium tuberculosis* isolates in Songjiang, Shanghai. We used whole-genome sequencing to discriminate apparent genetic clusters of *M tuberculosis* sharing identical variable-number-tandem-repeat (VNTR) patterns, and analysed the relations between proximity of residence and the risk of genomically clustered *M tuberculosis*. Finally, we used genomic, spatial, and epidemiological data to estimate time of infection and transmission links among migrants and residents.

Findings Between Jan 1, 2009, and Dec 31, 2015, 1620 cases of culture-positive tuberculosis were recorded, 1211 (75%) of which occurred among internal migrants. 150 (69%) of 218 people sharing identical VNTR patterns had isolates within ten single-nucleotide polymorphisms (SNPs) of at least one other strain, consistent with recent transmission of *M tuberculosis*. Pairs of strains collected from individuals living in close proximity were more likely to be genetically similar than those from individuals who lived far away—for every additional km of distance between patients' homes, the odds that genotypically matched strains were within ten SNPs of each other decreased by about 10% (OR 0·89 [95% CI 0·87–0·91]; p<0·0001). We inferred that transmission from residents to migrants occurs as commonly as transmission from migrants to residents, and we estimated that more than two-thirds of migrants in genomic clusters were infected locally after migration.

Interpretation The primary mechanism driving local incidence of tuberculosis in urban centres is local transmission between both migrants and residents. Combined analysis of epidemiological, genomic, and spatial data contributes to a richer understanding of local transmission dynamics and should inform the design of more effective interventions.

Funding National Natural Science Foundation of China, National Science and Technology Major Project of China, and US National Institutes of Health.

Introduction

Tuberculosis hits hardest in poor and vulnerable populations, and migrants are particularly affected. China has the third largest tuberculosis epidemic worldwide, with an estimated 900 000 new cases and 52 000 tuberculosis-related deaths in 2016.[1,2] In the past two decades, migration within China (internal migration), primarily from rural to urban environments, has become very common. Most internal migrants in China are young adults who travel from rural areas to enter the wage economy in urban centres. In 2010, an estimated 260 million people migrated internally (18% of the total national population), a number that is expected to increase in the coming decade.[3] This mass internal migration presents new challenges for tuberculosis control in Chinese cities.[4]

In China, the notification rate of tuberculosis is substantially higher in rural than in urban communities, and increased notifications and transmission of tuberculosis in cities have been predicted as a result.[4,5]

This hypothesis is supported by the fact that rural-to-urban internal migrants face substantial restrictions in accessing subsidised housing, social security, and medical care and insurance in urban centres, at least partly because of China's household registration (*hukou*).[4,6–8] Migrants are also more likely than urban residents to share crowded living conditions, be of low socioeconomic status, and have minimal formal education, all of which are often associated with increased risk of tuberculosis.[7,8] Others have documented that such internal migrants have a greater delay between symptom onset and clinical presentation or tuberculosis diagnosis and have less favourable treatment outcomes than urban residents.[9–11]

We combined epidemiological, molecular genetic, and spatial analysis to investigate the transmission dynamics of tuberculosis in Shanghai, China, which has had large population increases because of internal migration during the past two decades (in 2010 migrants constituted greater than 40% of the city's population).

Evidence before this study

We searched PubMed with the key terms "tuberculosis", "China", "migrant", and "floating population" for articles published in English before Dec 31, 2017. This search identified 55 articles, two of which included information about the spatial distribution of migrants with tuberculosis in Beijing. Several studies and reviews have also highlighted the challenges of tuberculosis control associated with the rise in internal migration from rural areas with high tuberculosis burdens to urban areas with low tuberculosis burdens in China. These studies focused on risk factors for tuberculosis among migrants and delays in diagnosis and treatment of this vulnerable population. We also searched the CNKI and WANFANG databases with similar key terms for articles published in Chinese before Dec 31, 2017. We identified two papers in Chinese describing the spatial distributions of tuberculosis cases in Songjiang on the basis of spatial scan statistics. We found no studies of how rural-to-urban population shifts in China contribute to tuberculosis dynamics in large cities.

Added value of this study

We combined genomic, geospatial, and epidemiological analyses to understand the transmission dynamics of tuberculosis in Songjiang, a district in Shanghai with large population influxes from rural China. We showed that increases in internal migration have been associated with increasing overall rates of tuberculosis notification in Songjiang, and that migrants have significantly higher rates of tuberculosis than residents. Our results strongly suggested that most clustered disease in migrants was the result of local transmission events within Songjiang, and that transmission from residents to migrants occurred as commonly as transmission from migrants to residents.

Implications of all the available evidence

Our finding that increasing tuberculosis notification rates in Songjiang were not attributable only to importation of tuberculosis from rural areas, but rather reflected complex patterns of local transmission, suggests the need for new interventions that can interrupt transmission in cities (eg, introduction of more active approaches for case finding and reduction of barriers to diagnosis and treatment for both migrants and residents). Our study also showed that the combination of genomic, spatial, and epidemiological data can be used to characterise the dynamics of tuberculosis in ways that were not previously possible.

Methods

Study design and participants

We did a population-based study of all culture-positive tuberculosis cases diagnosed in residents of the Songjiang district of Shanghai, China. Most internal migrants have settled in seven suburban districts in Shanghai, and Songjiang is typical in terms of the fraction of the population that migrants comprise, the demographic, educational, and job-seeking characteristics of migrants, and the fraction of notified tuberculosis that occurs among migrants (appendix). Roughly 1·10 million (62%) of Songjiang's 1·77 million residents are internal migrants, whom we defined as people without a Shanghai household registration status through the Chinese *hukou* system. Beginning in 2004, the Shanghai Center for Disease Control and Prevention (CDC) implemented a new policy extending free tuberculosis treatment to all migrants (previously free treatment was not available outside individuals' originally registered residence).

Our study included all people aged 15 years or older with tuberculosis symptoms, including cough for at least 2 weeks, fever, chest pain, weight loss, night sweats, and abnormal chest radiographs. Patients in prison were excluded. Community physicians routinely identified and referred individuals with symptoms to hospitals designated for tuberculosis diagnosis in Songjiang. All community health-care centres were included. All included individuals provided written informed consent. The study was reviewed and approved by the Shanghai CDC and Institute of Biomedical Sciences of Fudan University (appendix).

Procedures

Three sputum samples (spot, early morning, and night) were collected from people with tuberculosis symptoms. The samples were tested for *Mycobacterium tuberculosis* by microscopy and culture, consistent with national guidelines.[2] Sputum induction was used when people could not produce sputum spontaneously. One pre-treatment sputum specimen was submitted to the Shanghai CDC tuberculosis reference laboratory for each person with culture-positive tuberculosis, and rifampicin and isoniazid drug-susceptibility testing was done by the proportion method in Löwenstein-Jensen media.

Demographic, clinical, and microbiological records were obtained from the routine tuberculosis surveillance system.[12] Epidemiological investigations were done within a week of tuberculosis diagnosis with a standardised questionnaire administered by local community Shanghai CDC staff at community health-care centres. Information collected included current residential address, original household registration province and home address of migrants, identities of close contacts (ie, family members, work colleagues, and classmates), known history of exposure to individuals with active tuberculosis, and specific locations visited where transmission could plausibly have occurred.

Genotyping and whole-genome sequencing

We used the cetyl trimethyl ammonium bromide method to extract genomic DNA from cultures from one sputum specimen per patient. Beijing family strains (ie, the most

A
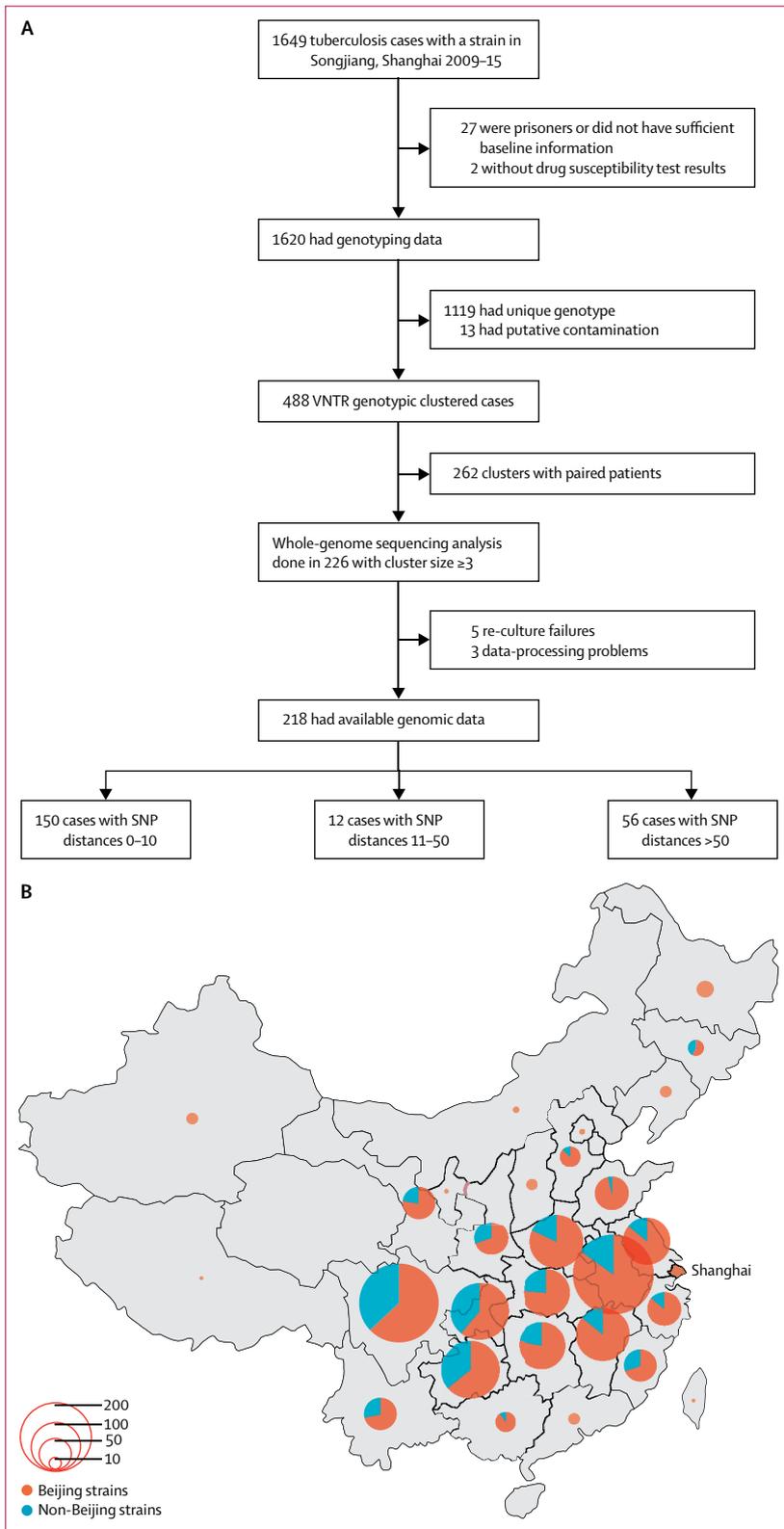
prevalent strains in China) were identified by a PCR-based assay and were confirmed by examination of specific single-nucleotide polymorphisms (SNPs).[13] We first did variable number tandem repeat (VNTR) genotyping for each strain, with a 9+3 loci set with hypervariable loci that was developed specifically for discriminating strains within China.[14]

To further discriminate closely related strains, we did whole-genome sequencing for the subset of strains that shared identical VNTR genotypes.[15] Because of resource limitations, we focused sequencing efforts on the subset of strains that were in VNTR clusters of three or larger. Whole-genome sequencing was done with the Illumina Hiseq 2000 (Illumina, San Diego, CA, USA) with an expected coverage of 100×. Paired-end reads were mapped to the reference genome H37Rv (GenBank AL123456) with Bowtie 2 (version 2.3.1). The SAMtools (version 1.6) and VarScan (version 2.3.6) suite were used to define SNPs (frequency ≥75; appendix). Genes associated with drug resistance and *PE-PGRS* and *PE-PPE* genes were excluded from our SNP analysis.[16]

For our primary analysis, we defined genomically clustered strains as those within a threshold of ten or fewer SNPs.[17] We varied this threshold in subsequent sensitivity analyses. We examined the genetic distance among pairs of *M tuberculosis* isolates from three different VNTR cluster types: those consisting of residents only, those consisting of migrants only, and those consisting of both residents and migrants (mixed clusters). The sequencing data were deposited in the National Center for Biotechnology Information Sequence Read Archive (SRP124760).

Spatial distance analysis and transmission inference

The residential address of each patient with tuberculosis at the time of diagnosis was geocoded with ArcGIS (Esri, Redlands, CA, USA) and Google Maps tools (Google, Mountain View, CA, USA). We also used a Chinese-language-based web geocoding tool from Baidu Maps (Baidu, Beijing, China) to verify locations. We used kernel density estimation to identify patterns of spatial aggregation of patients with tuberculosis.[18] We used multivariable logistic regression to test the hypothesis that spatial proximity was associated with greater genetic similarity (ie, a smaller SNP distance) among pairs of *M tuberculosis* isolates (appendix). The model included adjustment for genetic sub-lineage of the strains and an indicator variable, which tracked whether the pair of isolates was from two migrants, two residents, or one resident and one migrant.

Figure 1: **Study flowchart (A) and distribution of provinces of origin for migrant patients with tuberculosis in Songjiang, Shanghai, 2009–15 (B)**
VNTR=variable number of tandem repeats. SNP=single-nucleotide polymorphism. The size of the circles in (B) represents the number of culture-positive tuberculosis cases, stratified by Beijing strains (orange) and non-Beijing strains (blue).

We reconstructed plausible transmission networks on the basis of phylogenetic analysis (based on the median-joining method with NETWORK [version 5.0]), epidemiological linkages from questionnaire data,[19] and spatial distribution of the residential locations of the genomically clustered patients with tuberculosis. For ease of visualisation, we focused on clusters of at least four individuals. We also used a Bayesian evolutionary analysis to infer a time-labelled phylogeny by sampling trees with BEAST (version 1.8.4), and subsequently used a Markov chain Monte Carlo approach (with TransPhylo) in the R package (version 3.3.2) to infer the estimated infection times and putative transmission directions between resident and migrant patients on the basis of a consensus transmission tree that summarised the outputs (appendix).[20] We quantified the relative importance of importation of latent or incipient tuberculosis with the importance of local transmission by comparing the estimated time of transmission with reported time of arrival in Songjiang.

Role of the funding source

The study funders had no role in study design; data collection, analysis, or interpretation; or writing of the report. The corresponding author had full access to all study data and final responsibility for the decision to submit for publication.

Results

From Jan 1, 2009, to Dec 31, 2015, 1649 people were diagnosed with culture-positive tuberculosis in the Songjiang district. 29 people were excluded because they were incarcerated prisoners, because sufficient baseline information was not available, or because drug-susceptibility testing was not done (figure 1A). Of the remaining 1620 individuals with tuberculosis, 409 (25%) were residents and 1211 (75%) were internal migrants. 1073 (89%) of migrants with incident tuberculosis came to Songjiang from central and western regions of China, both of which have substantially higher tuberculosis notification rates than Shanghai (figure 1B; appendix). 776 of 1031 migrants migrated to Songjiang less than 5 years before being diagnosed with tuberculosis (appendix). The remaining 180 migrants did not provide this information (appendix).

Migrants with tuberculosis were significantly younger (median age 27 years *vs* 55 years; p<0·0001), were more likely to be female (36% *vs* 21%; p<0·0001), and less likely to have been previously treated for tuberculosis (7% *vs* 13%; p=0·0002) than residents with tuberculosis (table 1). Migrants with tuberculosis were less likely to have been notified and diagnosed in specialist tuberculosis hospitals than resident patients (14% *vs* 25%; p<0·0001).

The appendix shows crude notification rates for culture-positive tuberculosis by age. Based on population data from the 2010 national census,[21] the age-adjusted notification rate of culture-positive tuberculosis in migrants was higher than that in residents (18·20 per 100 000 per year *vs* 8·35 per

	Migrant patients (n=1211)	Resident patients (n=409)	p value*
Sex			
Female	441 (36%)	87 (21%)	<0·0001
Male	770 (64%)	322 (79%)	Reference
Age, years			
Median (IQR)	27 (22–38)	55 (37–71)	<0·0001
15–24	471 (39%)	48 (12%)	Reference
25–44	567 (47%)	83 (20%)	0·05
45–64	150 (12%)	130 (32%)	<0·0001
>65	23 (2%)	148 (36%)	<0·0001
Occupation			
Commercial service	245 (20%)	70 (17%)	Reference
Physical labour	755 (62%)	85 (21%)	<0·0001
Farmer	37 (3%)	115 (28%)	<0·0001
Student	45 (4%)	27 (7%)	0·007
Retired	10 (1%)	62 (15%)	<0·0001
Other	119 (10%)	50 (12%)	0·09
Case detection			
Referral	610 (50%)	160 (39%)	Reference
Follow-up after referral	347 (29%)	111 (27%)	0·15
Clinical consultation (for tuberculosis)	166 (14%)	104 (25%)	<0·0001
Physical examination	84 (7%)	31 (8%)	0·13
Other	4 (<1%)	3 (1%)	0·15
Previously treated for tuberculosis	87 (7%)	54 (13%)	0·0002
Median diagnostic delay, days (IQR)	26 (15–47)	30 (18–46)	0·45
Cavitary tuberculosis	509 (42%)	180 (44%)	0·48
Treatment outcomes			
Cured or completed treatment	1006 (83%)	317 (78%)	Reference
Treatment unsuccessful	9 (1%)	4 (1%)	0·56
Death	8 (1%)	18 (4%)	<0·0001
Lost to follow-up	7 (1%)	0 (0%)	0·36
On treatment	156 (13%)	56 (14%)	0·43
Other	25 (2%)	14 (3%)	0·08
Laboratory testing			
Beijing strains	928 (77%)	359 (88%)	<0·0001
Sputum smear positive	762 (63%)	264 (65%)	0·53
Drug-susceptible	1072 (89%)	356 (87%)	Reference
Mono-drug-resistant	92 (8%)	36 (9%)	0·42
Multidrug-resistant†	47 (4%)	17 (4%)	0·81

Data are n (%) or median (IQR). *p values are for comparisons between migrant and resident patients. †Resistant to at least isoniazid and rifampicin.

Table 1: Characteristics of migrant and resident patients in Songjiang, Shanghai, 2009–15

100 000 per year). Of the 1607 culture-positive individuals for whom sufficient genotyping data were available, isolates from 488 (30%) shared identical VNTR patterns (figure 1A). These isolates were grouped in 185 VNTR clusters of sizes ranging from two (131 clusters) to 11 (one cluster) patients (appendix). Beijing family strains were the most commonly observed isolates (1274 [79%] of 1607).

We obtained sufficient whole-genome-sequencing data from 218 of 226 isolates from 52 different VNTR clusters

(figure 1A). Analysis of pairwise nucleotide differences within each VNTR cluster showed that isolates from 150 individuals (102 migrants and 48 residents) were within ten SNPs of at least one other isolate from the same VNTR cluster, resulting in 44 genomic clusters of

Figure 2: Distribution of pair-wise genetic distances (number of SNPs) within VNTR-based clusters
SNP=single-nucleotide polymorphism. VNTR=variable number of tandem repeats. The dashed lines show the ten SNP threshold.

two-to-eight patients (appendix). After exclusion of the pairs separated by more than 100 SNPs (which were clearly not related through recent transmission by any reasonable SNP threshold), isolates from resident-only VNTR clusters had a significantly larger genetic distance (median difference 19 SNPs) than migrant-only (median difference two SNPs) and mixed VNTR clusters (median difference three SNPs; $p < 0.0001$; figure 2).

Although the patterns of genotypic and genomic clustering suggest the possibility of local transmission between migrants and between migrants and residents within Songjiang, migrants could have been infected before their arrival, and thus such genetic similarity might be an artifact of infections imported from other parts of the country. Further inspection of genomic clusters including at least two migrants showed that 95 (84%) of 113 migrants in such genomic clusters travelled to Songjiang from different provinces, suggesting that such clustering probably reflects local transmission after arrival in Songjiang (appendix).

Migrants in Songjiang resided primarily within several sub-districts with industrial parks (appendix). One of these industrial sub-districts had the highest tuberculosis notification rate overall (figure 3A); 254 (94%) of the 269 notifications in this sub-district were migrants. Kernel density analysis of home residence showed two distinct spatial distributions, with tuberculosis among migrants concentrated in several areas of eastern Songjiang (figure 3B) and among residents concentrated in the centre of the district (ie, downtown; figure 3C)—a pattern that was consistent with the overall distribution of migrants and residents in Songjiang (appendix).
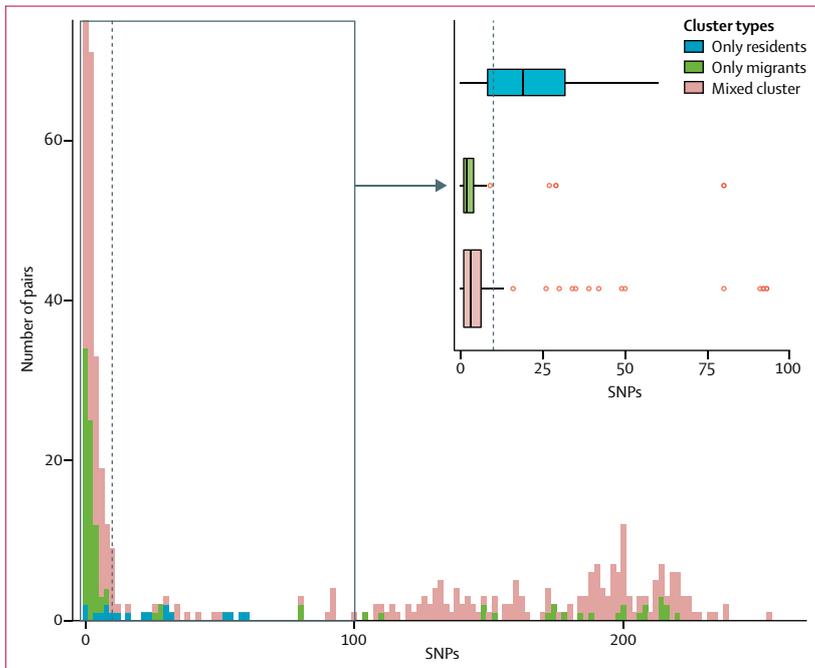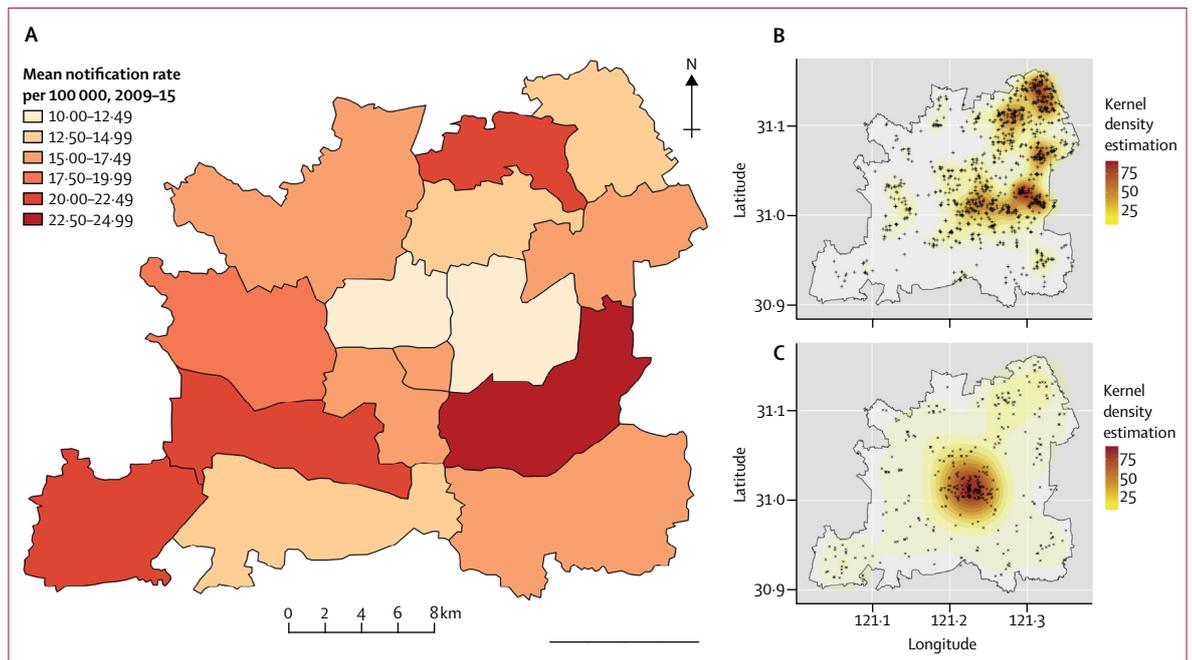

Figure 3: Overall culture-positive tuberculosis notification rate by sub-district (A), and kernel density estimation for migrants (B) and residents (C) in Songjiang, Shanghai, 2009–15

Table 2 shows the results of the multivariable logistic regression; a second unadjusted model had similar results, but higher Akaike information criterion values. The odds of two strains being genomically clustered decreased by about 10% for every additional km between residences; this result was robust with different thresholds for definition of the strains as genetically clustered (table 2). Median geographical distance between pairs of isolates increased as the genetic distance threshold for a genomic cluster increased (appendix).

Epidemiological data (which were available before pathogen genetic data) suggested possible links among only 26 (18%) of the 142 individuals within genomic clusters. Four (9%) of the 46 residents in genomic clusters reported epidemiological links to other cases (all among reported close contacts), compared with 22 (23%) of the 96 migrants (p=0·04). No links between migrants and residents were identified through routine epidemiological investigations.

Putative transmission networks among individuals sharing genomically clustered strains showed that, despite the absence of epidemiological links, genetic distances between a migrant and resident in any of the mixed clusters were no greater than a single SNP (figure 4). Three genomic clusters suggested that transmission of multidrug-resistant tuberculosis between residents and migrants occurred (figure 4A; appendix). We noted spatial aggregation within most genomic clusters, including in many migrant-only clusters (11, 12, 13, and 15; figure 4B), one resident-only cluster (02), and all the mixed clusters except cluster 10. In some of the mixed clusters, residents did not live close to migrants but lived in proximity to one another, as did the migrants—eg, mixed clusters 03 and 09, in which all the residents lived close to one another in the downtown area, whereas the migrants lived in industrial areas elsewhere in the district. We noted similar patterns of spatial aggregation when considering genomic clusters of isolates from three individuals (appendix).

After exclusion of the initial transmission event leading to the putative source case of each cluster, we estimated 74 transmission events occurring among 72 genomically clustered patients, 37 of which were captured during this study (appendix). We did not find evidence that migrants were more likely to transmit to residents than residents were to migrants: 12 (55%) of 22 estimated transmission events initiated by a migrant resulted in a secondary case in a resident, and nine (60%) of 15 estimated transmission events initiated by a resident resulted in a secondary case in a migrant (p=0·7). 47 (78%) of the 60 migrants with sufficient data for this analysis were probably infected after their arrival in Songjiang (appendix)—consistent with the finding that migrants sharing genomically clustered strains usually originated from different home provinces (appendix).

Discussion

In this 7-year population-based study involving whole-genome sequencing, geospatial analysis, and epidem-

	Distance between homes (per each additional km)		p value	Akaike information criterion
	Odds ratio*	95% CI		
3	0·90	0·87–0·93	<0·0001	1516
5	0·89	0·87–0·92	<0·0001	1792
10	0·89	0·87–0·91	<0·0001	2020
50	0·90	0·87–0·92	<0·0001	2189
100	0·93	0·91–0·95	<0·0001	3188

SNP=single-nucleotide polymorphism. *Adjusted for covariates including household registration status (ie, migrant or resident) and genetic sub-lineages.

Table 2: Logistic regression of association between spatial distance and genetic relatedness, by genetic distance of genomic cluster (SNPs)

iological investigation of tuberculosis in Songjiang, Shanghai, immigration from rural areas was associated with a substantial increase in local tuberculosis notifications. The increase in the local burden of tuberculosis was not attributable only to importation and reactivation of latent infections acquired elsewhere. Rather, local transmission in Songjiang, occurring among migrants and between residents and migrants, was the dominant underlying mechanism driving increasing case notifications in this urban setting. These findings suggest both opportunities and challenges for improvement of tuberculosis control in cities.

The influx of internal migrants from rural areas has been suggested as an important driver of the increasing tuberculosis burden in urban settings in China,[22] particularly in populous cities like Beijing, Guangzhou Shenzhen, and Shanghai, where 40–70% of the total population consists of the internal migrants.[8,23–26] In our study, migrants comprised 75% of all culture-positive tuberculosis cases in Songjiang and had substantially higher crude and age-adjusted tuberculosis notification rates than residents. Administrative and socioeconomic barriers,[7] and challenging working and living conditions characterised by overcrowding and poor sanitation,[27] probably contribute to this increased risk in migrants.[7] Similar trends of rural-to-urban migration and fractions of disease among migrants[7,8,23–26] suggest that our findings could be generalisable to other cities in China.

By using genomic, epidemiological, and spatial data, we showed that transmission of tuberculosis occurs not only among migrants, but also between residents and migrants in Songjiang. Traditional epidemiological investigations did not identify most estimated trans-mission links, and were especially poor for identification of links between migrants and residents. We uncovered previously unsuspected links, and showed that genomically clustered strains infecting both migrants and residents were as likely to have been transmitted by residents to migrants as from migrants to residents. We inferred that most recent *M tuberculosis* infections that led to disease among migrants occurred after arrival in Songjiang. This finding
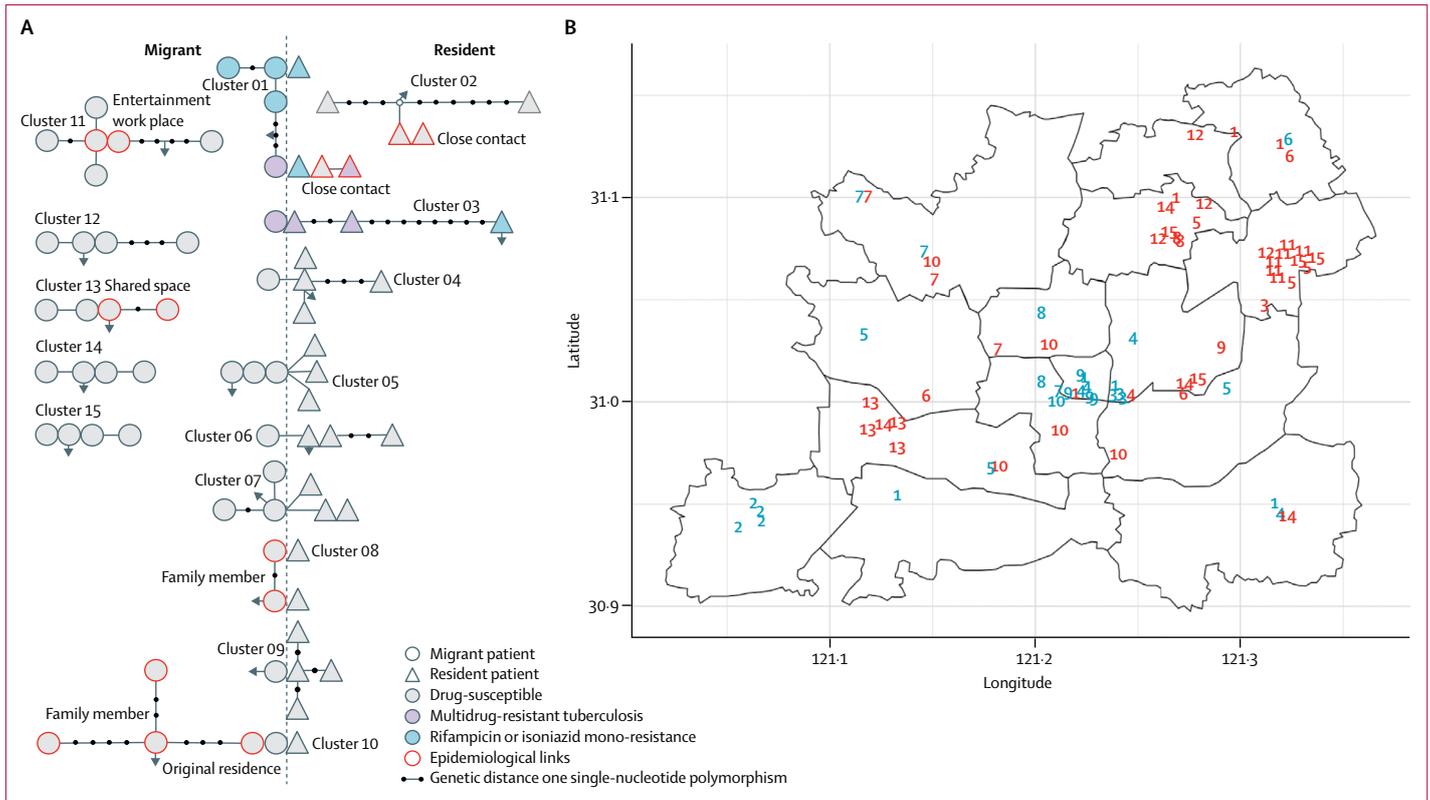
Figure 4: Genetic clusters with at least four patients (A), and spatial distribution of genetic clusters (B)

In (A), *Mycobacterium tuberculosis* isolates are separated by lines with length (as suggested by the dots) representing genetic distance. Arrows show the next closest isolate in the sequenced collection. In (B), orange numbers represent migrant patients and blue numbers represent resident patients.

emphasises that the increasing case notifications in Songjiang are not attributable only to importation of incipient tuberculosis from rural areas. From a public health perspective, this finding is notable because it suggests that emphasis should be placed on interventions that can interrupt transmission within Songjiang.

We detected a significant association between geographical proximity of residences of people with tuberculosis and the genetic relatedness of the *M tuberculosis* strains causing their disease. Although previous studies have shown that geographical aggregation of tuberculosis cases can be associated with genotypic clustering,[28–30] the relatedness of spatial proximity and genetic distance of *M tuberculosis* transmission was not examined in any of them. We also showed a positive association between spatial proximity and genetic relatedness for almost all genomic clusters of at least three people, some of which suggested extreme geographical proximity (eg, cluster 11, which consisting of six migrants, all of whom lived near a single industrial park).

The identification of spatial aggregation of closely related strains among migrant-only and resident-only clusters suggests contacts occurring in shared local social networks.[27] For mixed clusters, we identified more dispersed spatial patterns, because migrants and residents tend to live in distinct neighbourhoods. Transmission

events between migrants and residents suggest that transmission could occur through more casual contact in areas where populations intermix, such as the downtown area, which serves as a centre for transportation and entertainment.

Our study has noteworthy limitations. First, and most importantly, our ability to reliably infer patterns of tuberculosis transmission in Songjiang was dependent on how complete our sampling of *M tuberculosis* isolates was during the study. We know that not all notified cases had culture-positive disease, and we could not include people without isolates available for genotyping and sequencing in our analyses. Furthermore, we probably missed cases occurring among migrants who returned to their home provinces for diagnosis or treatment, and the potential bias associated with these missing cases is difficult to quantify. However, the fact that migrants had better health-care access and resources in Shanghai than in other cities in China during our study probably minimised the effect of these missing cases. The completeness of sampling of *M tuberculosis* isolates also affected the quality of our inference about transmission and estimation of most probable times of infection. Second, we could not estimate the overall proportion of strains that were genomically clustered (eg, within a particular SNP distance from

another strain) because whole-genome sequencing was restricted to VNTR clusters of three or larger. Finally, our spatial analysis was restricted to residential addresses at the time of diagnosis. Although the routine investigation included questions about frequently visited locations, this information was not uniformly collected and we did not attempt to use it in our analysis beyond noting when several cases in a genomic cluster reporting a shared location.

In conclusion, genomic, spatial, and epidemiological data helped us to better understand the complex dynamics of tuberculosis transmission in an urban centre that is undergoing important changes in migration. That local transmission not attributable to imported disease was an important driver of infection should motivate tuberculosis control programmes to increase access to diagnosis and care and encourage more intensive case-finding activities to halt tuberculosis transmission among both migrants and residents.

References
1 WHO. Global tuberculosis report 2017. http://www.who.int/tb/ publications/global_report/en/ (accessed Aug 2, 2017).
2 Wang L, Liu J, Chin DP. Progress in tuberculosis control and the evolving public-health system in China. *Lancet* 2007; **369**: 691–96.
3 Peng X. China's demographic history and future challenges. *Science* 2011; **333**: 581–87.
4 Tobe RG, Xu L, Song P, Huang Y. The rural-to-urban migrant population in China: gloomy prospects for tuberculosis control. *Biosci Trends* 2011; **5**: 226–30.
5 Dhavan P, Dias HM, Creswell J, Weil D. An overview of tuberculosis and migration. *Int J Tuberc Lung Dis* 2017; **21**: 610–23.
6 Wei X, Chen J, Chen P, et al. Barriers to TB care for rural-to-urban migrant TB patients in Shanghai: a qualitative study. *Trop Med Int Health* 2009; **14**: 754–60.
7 Hu X, Cook S, Salazar MA. Internal migration and health in China. *Lancet* 2008; **372**: 1717–19.
8 Jia ZW, Jia XW, Liu YX, et al. Spatial analysis of tuberculosis cases in migrants and permanent residents, Beijing, 2000–2006. *Emerg Infect Dis* 2008; **14**: 1413–19.
9 Zhou C, Tobe RG, Chu J, Gen H, Wang X, Xu L. Detection delay of pulmonary tuberculosis patients among migrants in China: a cross-sectional study. *Int J Tuberc Lung Dis* 2012; **16**: 1630–36.
10 Wang W, Jiang Q, Abdullah AS, Xu B. Barriers in accessing to tuberculosis care among non-residents in Shanghai: a descriptive study of delays in diagnosis. *Eur J Public Health* 2007; **17**: 419–23.
11 Tobe RG, Xu L, Zhou C, Yuan Q, Geng H, Wang X. Factors affecting patient delay of diagnosis and completion of direct observation therapy, short-course (DOTS) among the migrant population in Shandong, China. *Biosci Trends* 2013; **7**: 122–28.
12 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**: e51133.
13 Yang C, Luo T, Sun G, et al. *Mycobacterium tuberculosis* Beijing strains favor transmission but not drug resistance in China. *Clin Infect Dis* 2012; **55**: 1179–87.
14 Yang C, Shen X, Peng Y, et al. Transmission of *Mycobacterium tuberculosis* in China: a population-based molecular epidemiologic study. *Clin Infect Dis* 2015; **61**: 219–27.
15 Luo T, Yang C, Peng Y, et al. Whole-genome sequencing to detect recent transmission of *Mycobacterium tuberculosis* in settings with a high burden of tuberculosis. *Tuberculosis (Edinb)* 2014; **94**: 434–40.
16 Casali N, Broda A, Harris SR, Parkhill J, Brown T, Drobniewski F. Whole genome sequence analysis of a large isoniazid-resistant tuberculosis outbreak in London: a retrospective observational study. *PLoS Med* 2016; **13**: e1002137.
17 Nikolayevskyy V, Kranzer K, Niemann S, Drobniewski F. Whole genome sequencing of *Mycobacterium tuberculosis* for detection of recent transmission and tracing outbreaks: a systematic review. *Tuberculosis (Edinb)* 2016; **98**: 77–85.
18 Ribeiro FK, Pan W, Bertolde A, et al. Genotypic and spatial analysis of *Mycobacterium tuberculosis* transmission in a high-incidence urban setting. *Clin Infect Dis* 2015; **61**: 758–66.
19 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**: 275–84.
20 Didelot X, Gardy J, Colijn C. Bayesian inference of infectious disease transmission from whole-genome sequence data. *Mol Biol Evol* 2014; **31**: 1869–79.
21 National Bureau of Statistics of China. The 2010 population census of the People's Republic of China. http://www.stats.gov.cn/english/ statisticaldata/censusdata/ (accessed April 23, 2013).
22 Sun YX, Zhu L, Lu ZH, Jia ZW. Notification rate of tuberculosis among migrants in China 2005-2014: a systematic review and meta-analysis. *Chin Med J* 2016; **129**: 1856–60.
23 Zhang LX, Tu DH, An YS, Enarson DA. The impact of migrants on the epidemiology of tuberculosis in Beijing, China. *Int J Tuberc Lung Dis* 2006; **10**: 959–62.
24 Li X, Yang Q, Feng B, et al. Tuberculosis infection in rural labor migrants in Shenzhen, China: emerging challenge to tuberculosis control during urbanization. *Sci Rep* 2017; **7**: 4457.
25 Zeng J, Shi L, Zou X, Chen W, Ling L. Rural-to-urban migrants' experiences with primary care under different types of medical institutions in Guangzhou, China. *PLoS One* 2015; **10**: e0140922.
26 Chen J, Qi L, Xia Z, et al. Which urban migrants default from tuberculosis treatment in Shanghai, China? *PLoS One* 2013; **8**: e81351.
27 Shen J, Huang Y. The working and living space of the 'floating population' in China. *Asia Pac Viewpoint* 2003; **44**: 11.
28 Smith CM, Maguire H, Anderson C, Macdonald N, Hayward AC. Multiple large clusters of tuberculosis in London: a cross-sectional analysis of molecular and spatial data. *ERJ Open Res* 2017; **3**.
29 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**: 287–94.
30 Bishai WR, Graham NM, Harrington S, et al. Molecular and geographic patterns of tuberculosis transmission after 15 years of directly observed therapy. *JAMA* 1998; **280**: 1679–84.