Social Mixing and Clinical Features Linked With Transmission in a Network of Extensively Drug-resistant Tuberculosis Cases in KwaZulu-Natal, South Africa

Kristin N. Nelson[1], Samuel M. Jenness[1], Barun Mathema[2], Benjamin A. Lopman[1], Sara C. Auld[1,3], N. Sarita Shah[4], James C.M. Brust[5], Nazir Ismail[6,7], Shaheed Vally Omar[6], Tyler S. Brown[8], Salim Allana[1], Angie Campbell[1], Pravi Moodley[9,10], Koleka Mlisana[9,10], Neel R. Gandhi[1,3]

[1] Rollins School of Public Health, Emory University, Atlanta, GA, USA

[2] Mailman School of Public Health, Columbia University, New York, NY, USA

[3] Emory University School of Medicine, Atlanta, GA, USA

[4] U.S. Centers for Disease Control and Prevention, Atlanta, GA, USA

[5] Albert Einstein College of Medicine and Montefiore Medical Center, Bronx, NY, USA

[6] National Institute for Communicable Diseases, Johannesburg, South Africa

[7] University of Pretoria, South Africa

[8] Massachusetts General Hospital, Infectious Diseases Division, Boston, MA, USA

[9] National Health Laboratory Service, South Africa

[10] School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa

Summary: We identified factors that lead to transmission of XDR TB, including contact with urban areas; these factors can suggest settings instrumental in transmission and indicate optimal locations and groups to target with interventions.

Running title: Factors linked with XDR TB transmission

Corresponding Author:

Neel R. Gandhi, MD

Phone: +1 404 727-2317; Fax: (404) 727-8737; Email: neel.r.gandhi@emory.edu

Abstract:

Background: Tuberculosis (TB) is the leading infectious cause of death globally and drug-resistant TB strains pose a serious threat to controlling the global TB epidemic. The clinical features, locations, and social factors driving transmission in settings with a high incidence of drug-resistant TB are poorly understood.

Methods: We measured a network of genomic links using *Mycobacterium tuberculosis* (*Mtb*) whole genome sequences.

Results: Cases with 2–3 months of cough or who spent time in urban locations, were more likely to be linked in the network, while cases with sputum smear-positive disease were less likely to be linked than those with smear-negative disease. Associations persisted using different thresholds to define genomic links and irrespective of assumptions about the direction of transmission.

Conclusions: Identifying factors that lead to many transmissions, including contact with urban areas, can suggest settings instrumental in transmission and indicate optimal locations and groups to target with interventions.

Keywords: tuberculosis, drug-resistance; extensively drug-resistant tuberculosis; molecular epidemiology; whole genome sequencing; transmission; exponential random graph models; network models; transmission networks

Introduction

Tuberculosis (TB) is the leading infectious cause of death globally and drug-resistant TB strains pose a serious threat to controlling the global TB epidemic[1]. The majority of drug-resistant TB cases in settings with a high TB incidence are due to transmission of drug-resistant strains, rather than acquisition of resistance through inadequate TB treatment[2, 3]. Interventions to reduce incidence must include efforts to reduce transmission, which will require a clear understanding of the clinical features, types of social contact, and locations driving the spread of TB[4, 5].

Transmission heterogeneity, or variation in the number of secondary infections caused by an index case, may play a critical role in shaping TB epidemiology[6-9]. On a population level, transmission heterogeneity arises as a result of three factors: the extent and duration of infectiousness, the susceptibility of exposed persons to infection and disease, and the rate of contact between infected and susceptible persons[10, 11]. The first two factors can be considered functions of the biologic features of disease; the third is related to social and behavioral patterns that define where and with whom infected persons spend time. Identifying the clinical features of index cases (biologic factors) that cause many secondary cases as well as the locations or types of contact (social mixing factors) driving transmission can identify opportunities for interventions. Control measures targeting individuals who are likely to generate many secondary cases disease, or implemented in specific locations of high transmission risk, will likely outperform measures targeting broader areas or populations[8, 12].

Extensively drug-resistant (XDR) TB is associated with poorer treatment outcomes than drug-susceptible or multidrug-resistant (MDR) TB, making prevention of transmission critical. However, the factors driving XDR-TB transmission in high-incidence settings remain poorly understood. Markers of high bacterial burden, including sputum smear status and cavitary disease on chest x-ray, characterize ‘classical’ infectious TB[13-15], but in high-incidence settings these clinical features alone cannot explain observed individual-level heterogeneity in TB transmission[6]. Given the role of coughing in expelling *Mycobacterium tuberculosis* (*Mtb*) from the respiratory tract and into the surrounding environment, duration of cough may also be associated with transmission. However, there are few studies that examine

cough duration as a predictor of transmission. An individual's likelihood of transmitting TB is also influenced by behaviors that enable person-to-person contact necessary for transmission. Although over 80% of transmission may occur outside of the home, there is limited understanding of the locations at which transmission occurs in community settings and among non-close contacts (i.e., persons who do not know one another)[16, 17].

Identifying factors associated with transmission events is challenging given that disease transmission is unobserved; however, the advent of bacterial whole genome sequencing (WGS) to characterize *Mtb* isolates provides new opportunities to study transmission. Relative to less sensitive genotyping methods, WGS allows for improved precision to identify cases with highly similar *Mtb* sequences and thereby generate hypotheses about transmission events. Such links can be used to create a network comprised of putative transmission links between cases, and analytic methods specific to networks provide a framework for studying case characteristics associated with transmission. Sequencing may prove especially useful in settings with a high incidence of TB, where the majority of transmission is expected to occur in community locations between individuals who may not know one another and would therefore not be identified through conventional contact investigation. HIV genome sequences have been used to construct transmission networks and characterize transmission clusters[18, 19]; we build on this by using network models to identify factors associated with transmission in a network based on sequencing data from TB cases[20, 21].

In this study, we used bacterial WGS to define plausible transmission links between XDR-TB cases in KwaZulu-Natal, South Africa. Using the resulting sequencing-based network, we identified clinical and social mixing factors associated with TB transmission. The broader goal of this study was to identify locations, types of contact, and clinical features linked with XDR-TB transmission in South Africa, providing evidence for the development of effective interventions targeted towards individuals and settings sustaining TB spread.

Methods

Study design and procedures

The Transmission of HIV-Associated XDR-TB (TRAX) study is a cross-sectional study that enrolled culture-confirmed pulmonary XDR-TB patients diagnosed from May 2011 through August 2014 in KwaZulu-Natal, South Africa[2]. The primary aim of the study was to estimate the proportion of XDR-TB cases that develop due to transmission versus acquired drug resistance resulting from inadequate treatment. Detailed methods of the TRAX study have been previously published[2]. Briefly, we identified XDR-TB cases through the referral laboratory that conducts drug-susceptibility testing for all public healthcare facilities in the province. All participants provided written informed consent.

We interviewed participants to collect demographic and clinical information, including previous TB disease and HIV status, which was confirmed using medical records. Participants completed a social network questionnaire that included information on close contacts, the location and duration of time spent in congregate locations, and hospital admissions in the five years prior to their XDR-TB diagnosis.

Measures

Sequencing and bioinformatics procedures have been described in detail elsewhere. Briefly, we obtained participants' diagnostic XDR-TB isolates, re-cultured on Löwenstein-Jensen slants and conducted population sweeps. Raw paired-end sequencing reads were generated on the Illumina (MiSeq) platform and aligned to the H37Rv reference genome (NC_000962.3). Single nucleotide polymorphisms (SNPs) were detected using standard pairwise resequencing techniques (Samtools v0.1.19) against the reference and filtered for quality, read consensus, and proximity to indels and PE/PPE gene regions. Alignment files can be found at NCBI Bioproject PRJNA476470.

We defined a directed genomic link as a pair of XDR-TB cases with 5 or fewer SNP differences between their *Mtb* sequences, a cutoff used by prior studies, and considered the case with the earliest diagnosis date as the primary case[22]. We constructed a network of directed genomic links between cases, in which each node in the network represented a diagnosed XDR-TB case and each link

represented a genomic link between two cases. For each case, we calculated the number of genomic links in which they were the primary (index) case.

The diagnostic XDR-TB sputum sample was used to determine smear status and grade. Cough duration was patient-reported at the time of enrollment. Chest x-ray results were abstracted from patient medical records at time of XDR-TB diagnosis, but 34% (116) of patients were missing x-ray results, so we included cavitary disease only in secondary analyses.

To define urban contact, we determined whether a case reported ever living in eThekwini (Durban), the primary urban center of the province, whether they had been admitted to a hospital there in the previous five years, or whether they reported spending time at a congregate location there. To measure hospital contact, we summed the total number of months patients reported spending in the hospital in the five years prior to diagnosis. Participants also completed a social network questionnaire, in which they reported the number of contacts with whom they spent more than two hours per week in the previous year. We used the number of contacts reported as a proxy measure of close contact.

Exponential random graph models

Conventional statistical models assume that the characteristics of each subject are independent, an assumption not met in disease transmission networks. In a transmission network, the unit of interest is a transmission link, which consists of two cases whose attributes may be correlated. ERGMs are a tool for statistically modeling the probability of a link to form between nodes (cases) in a network, accounting for the inherent correlation among attributes of cases within the network. We used ERGMs to express the probability that a transmission link occurs between two cases in the network as a function of demographic and clinical characteristics of each case. (Supp. Eq. 1)

We constructed parsimonious models, which excluded terms that did not change main effect estimates by more than 10% when removed from the model, and fully-specified models, which included additional terms for year of study enrollment and *Mtb* strain, both of which may be artifactually related to the number of genomic links per case.

In addition to the primary predictors of interest, we defined *a priori* and included potential confounders in models: sex, age, and HIV status. Code for all models is available at https://github.com/kbratnelson/tb-ergms.

Alternate models and sensitivity analyses

Given uncertainty regarding the SNP threshold for defining a direct transmission event, we fit models to networks constructed using two additional SNP thresholds to define genomic links: one more stringent (≤3SNPs) and one less stringent (≤10SNPs). We also tested a definition for genomic links that combined WGS with restriction fragment length polymorphism typing (RFLP) results, since conventional genotyping techniques like RFLP capture regions of the genome that sequencing does not. In addition to defining networks using directed genomic links, we tested associations in 'undirected' networks, in which we did not assume a direction of transmission.

Results

TRAX cohort

Between 2011 and 2014, we screened a convenience sample of 521 (51%) of 1027 culture-confirmed XDR-TB patients diagnosed in KwaZulu-Natal and enrolled 404 (78% of screened). Among the 404 participants, 234 (58%) were female, with a median age of 34 years (interquartile range [IQR]: 28–43). Three hundred eleven (77%) participants were HIV-positive. *Mtb* isolates from 344 (85%) participants passed all sequencing quality filters and were available for analysis (Table 1).

Table 1. TRAX cohort characteristics, n = 344

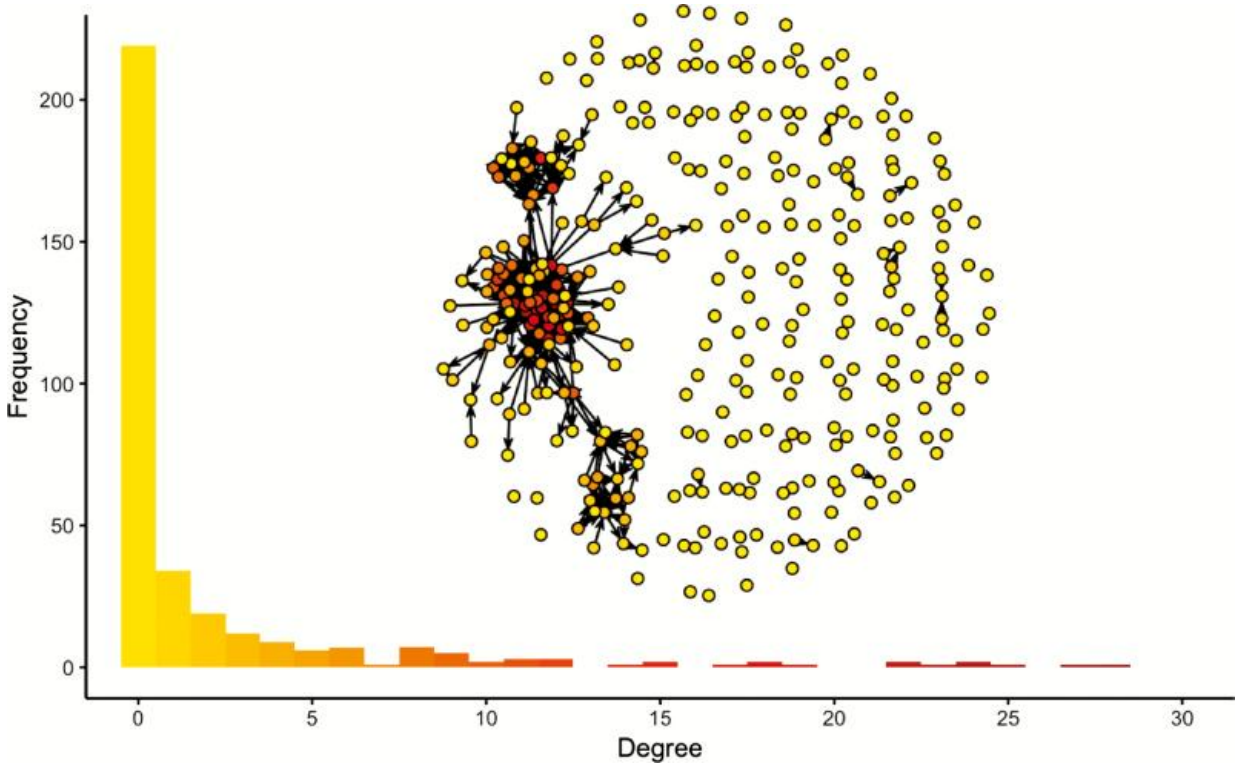
Characteristic	n (%), unless otherwise noted
Demographic	
Female	202 (59)
Age, median (IQR)	34 (29-43)
0-15	12 (3)
16-34	171 (50)
35-54	134 (39)
≥ 55	27 (8)
Monthly household income	
< R500	120 (35)
R500-R2,500	153 (44)
> R2,500	71 (21)
Clinical characteristics	
Current or former smoker	35 (10)
Diabetes	22 (6)
HIV positive	266 (77)
Receiving antiretroviral therapy	204 (77)
CD4 cell count (median, IQR)	240 (111-425)
Virologic suppression (<400 copies/mL)	134 (50)
Cough	
Patients with cough	284 (83)

Median duration of cough	8 (4-12)
Sputum smear negative for acid-fast bacilli	109 (32)
Scanty-positive	37 (11)
Smear-positive, grade 1	59 (17)
Smear-positive, grade 2	51 (15)
Smear-positive, grade 3 +	88 (26)
Chest x-ray available	228 (66)
Cavities on chest x-ray	60 (26)
Previous treatment for multidrug-resistant tuberculosis	105 (31)
Social mixing characteristics	
Number of reported contacts, median (IQR)	7 (4, 10)
Number of months in hospital, median (IQR)	3 (2, 5)
Previous stay at urban hospital	175 (51)
Current or previous urban residence	39 (11)
Reported visiting urban congregate setting	38 (11)

Genomic links

Among 344 cases comprising the sequencing-based network (threshold ≤5 SNPs), there were a total of 740 genomic links. 125 cases (36%) had at least one link. Among those cases with links, the number of links ranged from 1 to 28; 38 cases (30%) had between 1 and 5 links, 18 cases (14%) had more than 10 links (Figure 1). At a threshold of ≤10SNPs, there were 181 (53%) cases with at least one link; at ≤3SNPs, there were 116 (34%) cases with at least one link (S.Table 1).

Figure 1. Sequencing-based network, ≤ 5 SNP



Infectiousness measures

In our primary model, which adjusted for age, sex, and HIV status, cases reporting 2 or 3 months of cough were more likely to be linked than those reporting no cough. The odds of a genomic link among cases with 2 or 3 months of cough was 2.68 times (95%CI: 2.19, 3.27) higher and 2.31 times higher (95%CI: 1.91, 2.80) than those with no cough, respectively (Table 2). This trend did not continue in the highest category of cough duration (>4 months), but this group was small (n=33).

Table 2. Adjusted associations between clinical features and being linked in the directed network, ≤ 5 SNPs [a]

	n (%)	Odds Ratio	95% CI	p
Cough duration				
No cough reported	128 (37)	Ref	-	-
1 month	60 (17)	0.54	(0.39, 0.74)	< 0.01
2 months	51 (14)	2.68	(2.19, 3.27)	< 0.01
3 months	72 (21)	2.31	(1.91, 2.80)	< 0.01
≥ 4 months	33 (10)	1.06	(0.78, 1.45)	0.70
Smear status				
Smear -	109 (32)	Ref	-	-
Smear +, scanty +	37 (11)	0.98	(0.77, 1.25)	0.88
Smear +, grade 1	59 (17)	0.65	(0.52, 0.83)	< 0.01
Smear +, grade 2	51 (15)	0.73	(0.58, 0.91)	< 0.01
Smear +, grade 3+	88 (26)	0.58	(0.47, 0.72)	< 0.01

[a] Model also includes terms for HIV status, sex, and age.

In the same model, smear-positive cases were less likely to be linked than smear-negative cases, irrespective of smear grade. Cases with the highest smear grade of 3+ were the least likely to be linked (OR: 0.58, 95% CI: 0.47, 0.72) (Table 3); this association was similar in models excluding cough (S. Table 2). When considered as a dichotomous variable (smear-negative and smear-positive), smear-positive cases were 0.65 times as likely to have a link as smear-negative cases (95% CI: 0.55, 0.76) (S.Table 3). In models including only patients with available chest x-ray results, cavitary disease was associated with a higher likelihood of genomic links (S.Table 4).

Table 3. Adjusted associations between social mixing measures and being linked in the directed network [1], ≤ 5 SNPs

	n (%)	Odds Ratio	95% CI	p
Contact with urban areas				
0 urban locations	149 (43)	Ref	-	-
1 urban location	147 (43)	2.57	(2.17, 3.05)	< 0.01
≥ 2 urban locations	48 (20)	1.73	(1.33, 2.25)	< 0.01
Duration in hospital				
0 - 2 months	113 (33)	Ref	-	-
3 - 5 months	81 (24)	0.82	(0.69, 0.98)	0.03
> 5 months	59 (17)	0.38	(0.29, 0.50)	< 0.01
Named contacts				
0 - 4 contacts	108 (31)	Ref	-	-
5 - 10 contacts	144 (42)	1.17	(0.98, 1.41)	0.08
> 10 contacts	85 (25)	1.44	(1.18, 1.76)	< 0.01

[1] Model also includes terms for HIV status, sex, and age, smear status and grade, and cough duration.

In networks defined using a more stringent SNP threshold (≤3SNPs) and in which we did not assume a direction of transmission, the direction and magnitude of associations with infectiousness were generally similar (S.Table 5). In full models including terms for *Mtb* strain and year of study enrollment, results were similar to those from the primary model (S.Table 6).

Social mixing measures

In our primary model, cases reporting activity in 1 or more urban locations (residential, healthcare, or other congregate) were more likely to be linked than those who did not. (OR for 1 location: 2.57, 95%CI: 2.17, 3.05; OR for 2 or more locations: 1.73, 95%CI: 1.33, 2.25) (Table 3). When we deconstructed the variable describing contact with urban locations to determine if there was a specific component responsible for this association, we found that reporting a stay in an urban hospital was most strongly associated with being linked (OR: 2.65, 95%CI: 1.60, 4.39). However, cases who spent prolonged periods of time in the hospital (regardless of whether the hospital was in an urban location) were less likely to be linked in the network than those spending ≤2 months (OR for 3–5 months in hospital: 0.82, 95%CI: 0.69, 0.98; OR for 5 or more months: 0.38 , 95%CI: 0.29, 0.50). Number of named contacts was moderately associated with being linked in the network: cases who reported 5–10 or >10 contacts were more likely to be linked than cases who reported ≤4 named contacts (Table 3). Notably, associations with infectiousness measures persisted and were similar in magnitude in models including all social mixing variables. In models that also included terms for *Mtb* strain and year enrolled, associations were similar in direction and magnitude to results from the primary model (S.Table 7).

In networks with alternative thresholds for genomic links, the directions of associations were consistent with the primary model, though association strength was model-dependent (S.Table 8).

Discussion

We constructed a network of genomic links using *Mtb* whole genome sequences and found that XDR-TB cases reporting 2–3 months of cough or contact with urban locations were more likely to be linked. Contrary to the notion that smear positivity is a marker of infectiousness, we found that smear-positive cases were less likely to be linked than smear-negative cases. These associations persisted in networks using different thresholds to define genomic links.

Aside from the negative association with smear status, these findings are largely consistent with prior studies of TB transmission[13]. Cough duration and cavitary disease have previously been

associated with genotypic linkages in both drug-susceptible and multidrug-resistant TB[15, 23, 24]. However, these studies employed conventional genotyping methods; our study is the first to our knowledge to use WGS, which has higher specificity to detect transmission links. Although the trend between cough and genomic links did not persist in the group with the longest duration of cough symptoms, this may be due to the small size of this group, alternative etiologies of prolonged cough, or reflect true behavioral changes among cases with severe symptoms resulting in reduced contact rates and fewer transmission events.

There are several potential explanations for the unexpected finding that smear-positive cases were less likely to be linked than smear-negative cases. Patients with smear-negative disease may have less severe or clinically atypical disease, leading to delays in diagnosis, longer infectious periods and thus more opportunities for transmission. If smear-positivity does indeed indicate more severe disease, these cases may be more likely to die earlier due to unsuccessful treatment, or be more likely to be hospitalized, reducing opportunities for community transmission. Indeed, we found that the median time spent in hospital was longer among smear-positive than smear-negative cases (3 vs. 2 months). Lastly, we only considered a single sputum result at the time of diagnosis, which may fail to represent smear status over the course of TB disease.

Participants who spent time in urban locations prior to diagnosis were more likely to be linked in the network. Urban areas tend have higher incidence of TB and often provide ideal conditions for disease spread, as person-to-person contact rates may be higher than in rural areas. Moreover, urban areas of high TB incidence have been hypothesized to drive disease incidence in wider geographic areas, and our findings support previous findings from this same cohort suggesting that rural-urban migration may be driving transmission of XDR-TB in KwaZulu-Natal[25]. We found that the number of months spent in the hospital prior to diagnosis was negatively associated with being linked in the network. Healthcare facilities and other institutional settings are often considered to be ‘amplifiers’ of transmission; however, individuals who spent time in the hospital during their infectious periods may have lower effective

contact rates than those who spent their infectious periods living, working, and socializing in their communities[26].

We found a weak positive association with the number of contacts named by participants and being linked. We hypothesized that reporting more contacts would reflect a higher level of engagement in person-to-person contact (either through social activities, employment, school, or home life) and thus more opportunities for transmission. However, this trend did not persist among those reporting the highest numbers of contacts. This may reflect a threshold above which additional contact does not necessarily lead to additional transmission events, a notion supported by previous modeling studies, or that our measure of close contact is a poor proxy for the extent of true close contact during cases' infectious periods[6].

This study has several limitations. First, we enrolled 40% of all diagnosed XDR-TB cases in KwaZulu-Natal during this time period; therefore, there are missing cases and genomic links in this network. If missing cases are 'intermediaries' in the transmission chain between sampled cases we found to be linked, we might observe larger genomic differences between sampled cases than those that represent true transmission events. However, reducing the SNP threshold showed similar results to our primary analysis. Although cases enrolled in the study were demographically similar to all diagnosed XDR-TB cases, the extent to which enrolled cases are representative of unsampled, undiagnosed cases is not clear. In a modeling study examining the potential for biased sampling in this study, we have found that while cases with many transmission links may have been undersampled, our models do not suggest biased sampling by patient attributes like smear status. [27] Second, although we also conducted our analyses using a network in which we did not assume directionality and found largely similar results, we still cannot distinguish between individual-level factors that increase risk of infection from those that increase risk of transmission. Lastly, these transmission patterns may be different in settings with low TB incidence or low HIV prevalence, and further research should aim to characterize local TB transmission patterns in other contexts.

Identifying individual-level factors driving TB transmission can inform development of prevention strategies targeting specific groups to achieve maximal reductions in transmission. Our study suggests that cases who spend time in urban settings and those with extended symptom duration may cause many secondary TB cases, a finding which should be examined more thoroughly in future studies. Interventions targeting urban areas and those with high contact rates within them may be indicated.

Acknowledgments: We are grateful to the study team at the University of KwaZulu-Natal for their tireless efforts in data collection, record abstraction, participant recruitment, and interviews. We thank the participants and their families who consented to participate in this study.

Conflicts of Interest: The authors have no conflicts of interests to declare.

Disclaimer: The findings and conclusions in this manuscript are those of the authors and do not necessarily represent the official position of the funding agencies. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Financial Support: This study was primarily funded by grants from the US National Institute of Allergy and Infectious Diseases (NIAID), National Institutes of Health (NIH): R01AI089349 (PI Gandhi), R01AI087465 (PI Gandhi), and R01AI138646 (PI Gandhi). It was also supported in part by NIH/NIAID grants: R01AI114304 (PI Brust), K24AI114444 (PI Gandhi), K23AI134182 (PI Auld), Emory CFAR P30AI050409 (PI Curran), Einstein CFAR P30AI124414 (PI Goldstein), by Einstein/Montefiore ICTR UL1 TR001073 (PI Shamoon).

Contact Information: Correspondence and requests for reprints may be directed to Neel Gandhi (neel.r.gandhi@emory.edu) at 1518 Clifton Road NE, CNR 3031, Atlanta, GA 30322.

References

1. Global Tuberculosis Report. Geneva, Switzerland: World Health Organization, **2017**.
2. Shah NS, Auld SC, Brust JCM, et al. Transmission of Extensively Drug-Resistant Tuberculosis in South Africa. New England Journal of Medicine **2017**; 376:243-53.
3. 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. The Lancet Infectious Diseases **2017**; 17:275-84.
4. Auld SC, Kasmar AG, Dowdy DW, et al. Research Roadmap for Tuberculosis Transmission Science: Where Do We Go From Here and How Will We Know When We're There? The Journal of infectious diseases **2017**; 216:S662-s8.
5. The Paradigm Shift: Global Plan to End TB 2016-2020: Stop TB Partnership, **2015**.
6. McCreesh N, White RG. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. Scientific Reports **2018**; 8:5382.
7. Paull SH, Song S, McClure KM, Sackett LC, Kilpatrick AM, Johnson PTJ. From superspreaders to disease hotspots: linking transmission across hosts and space. Frontiers in ecology and the environment **2012**; 10:75-82.
8. Lloyd-Smith JO, Schreiber SJ, Kopp PE, Getz WM. Superspreading and the effect of individual variation on disease emergence. Nature **2005**; 438:355.
9. Li LM, Grassly NC, Fraser C. Quantifying Transmission Heterogeneity Using Both Pathogen Phylogenies and Incidence Time Series. Molecular Biology and Evolution **2017**; 34:2982-95.
10. VanderWaal KL, Ezenwa VO. Heterogeneity in pathogen transmission: mechanisms and methodology. Functional Ecology **2016**; 30:1606-22.
11. Dowdy DW, Azman AS, Kendall EA, Mathema B. Transforming the Fight Against Tuberculosis: Targeting Catalysts of Transmission. Clinical Infectious Diseases **2014**; 59:1123-9.
12. Dowdy DW, Golub JE, Chaisson RE, Saraceni V. Heterogeneity in tuberculosis transmission and the role of geographic hotspots in propagating epidemics. Proc Natl Acad Sci USA **2012**; 109.
13. Mathema B, Andrews JR, Cohen T, et al. Drivers of Tuberculosis Transmission. The Journal of infectious diseases **2017**; 216:S644-S53.
14. Behr MA, Warren SA, Salamon H, et al. Transmission of Mycobacterium tuberculosis from patients smear-negative for acid-fast bacilli. Lancet (London, England) **1999**; 353:444-9.
15. Rodwell TC, Kapasi AJ, Barnes RFW, Moser KS. Factors associated with genotype clustering of Mycobacterium tuberculosis isolates in an ethnically diverse region of southern California, United States. Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases **2012**; 12:1917-25.
16. Johnstone-Robertson SP, Mark D, Morrow C, et al. Social mixing patterns within a South African township community: implications for respiratory disease transmission and control. American journal of epidemiology **2011**; 174:1246-55.
17. Glynn JR, Guerra-Assuncao JA, Houben RMGJ, et al. Whole Genome Sequencing Shows a Low Proportion of Tuberculosis Disease Is Attributable to Known Close Contacts in Rural Malawi. PloS one **2015**; 10:e0132840.
18. Oster AM, Wertheim JO, Hernandez AL, Ocfemia MC, Saduvala N, Hall HI. Using Molecular HIV Surveillance Data to Understand Transmission Between Subpopulations in the United States. J Acquir Immune Defic Syndr **2015**; 70:444-51.
19. Morgan E, Oster AM, Townsell S, Peace D, Benbow N, Schneider JA. HIV-1 Infection and Transmission Networks of Younger People in Chicago, Illinois, 2005-2011. Public Health Reports **2017**; 132:48-55.
20. Klovdahl AS, Graviss EA, Yaganehdoost A, et al. Networks and tuberculosis: an undetected community outbreak involving public places. Social science & medicine (1982) **2001**; 52:681-94.
21. Oster AM, France A, Mermin J. Molecular epidemiology and the transformation of hiv prevention. Jama **2018**; 319:1657-8.

22. Hatherell H-A, Colijn C, Stagg HR, Jackson C, Winter JR, Abubakar I. Interpreting whole genome sequencing for investigating tuberculosis transmission: a systematic review. BMC medicine **2016**; 14:21.
23. Turner RD, Bothamley GH. Cough and the transmission of tuberculosis. The Journal of infectious diseases **2015**; 211:1367-72.
24. Turner RD HR, Birring SS, Bothamley Graham H. Daily Cough Frequency in Tuberculosis Is Associated with Rates of Household Infection. American Thoracic Society Conference, **2016**.
25. Nelson KN, Shah NS, Mathema B, et al. Spatial Patterns of Extensively Drug-Resistant Tuberculosis Transmission in KwaZulu-Natal, South Africa. The Journal of infectious diseases **2018**:jiy394-jiy.
26. Basu S, Stuckler D, McKee M. Addressing institutional amplifiers in the dynamics and control of tuberculosis epidemics. The American journal of tropical medicine and hygiene **2011**; 84:30-7.
27. Nelson KN, Gandhi NR, Mathema B, et al. Modeling missing cases and transmission links in networks of extensively drug-resistant tuberculosis in KwaZulu-Natal, South Africa. bioRxiv **2019**:655969.