nature communications

Article

https://doi.org/10.1038/s41467-024-52122-x

Mycobacterium tuberculosis cough aerosol culture status associates with host characteristics and inflammatory profiles

Received: 15 December 2023

Accepted: 27 August 2024

Published online: 01 September 2024

Check for updates

Videlis Nduba[1], Lilian N. Njagi[1], Wilfred Murithi[1], Zipporah Mwongera[1], Jodi Byers[2], Gisella Logioia[3], Glenna Peterson[3], R. Max Segnitz[3], Kevin Fennelly[4], Thomas R. Hawn[3] & David J. Horne[2,3]

Interrupting transmission events is critical to tuberculosis control. Cough-generated aerosol cultures predict tuberculosis transmission better than microbiological or clinical markers. We hypothesize that highly infectious individuals with pulmonary tuberculosis (positive for cough aerosol cultures) have elevated inflammatory markers and unique transcriptional profiles compared to less infectious individuals. We performed a prospective, longitudinal study using cough aerosol sampling system. We enrolled 142 participants with treatment-naïve pulmonary tuberculosis in Kenya and assessed the association of clinical, microbiologic, and immunologic characteristics with *Mycobacterium tuberculosis* aerosolization and transmission in 129 household members. Contacts of the forty-three aerosol culture-positive participants (30%) are more likely to have a positive interferon-gamma release assay (85% vs 53%, P = 0.006) and higher median IFNγ level (P < 0.001, 4.28 IU/ml (1.77-5.91) vs. 0.71 (0.01-3.56)) compared to aerosol culture-negative individuals. We find that higher bacillary burden, younger age, larger mean upper arm circumference, and host inflammatory profiles, including elevated serum C-reactive protein and lower plasma TNF levels, associate with positive cough aerosol cultures. Notably, we find pre-treatment whole blood transcriptional profiles associate with aerosol culture status, independent of bacillary load. These findings suggest that tuberculosis infectiousness is associated with epidemiologic characteristics and inflammatory signatures and that these features may identify highly infectious persons.

Tuberculosis (TB), a leading infectious disease-related cause of death[1], is spread person-to-person through the inhalation of aerosolized bacilli. A key step for TB elimination is interrupting transmission events to prevent new acquisition of infection and disease. However, knowledge gaps in understanding the biology and determinants of TB transmission undermine efforts to develop interventions[2,3]. These gaps include uncertainties around the determinants of infectiousness, poor estimations of individual infectiousness, and the lack of accurate and convenient biomarkers of infectiousness. Similar to other infectious diseases[4–6], "superspreaders", individuals with TB who are responsible for the majority of *Mycobacterium tuberculosis* (Mtb) transmission[7–11] may have an outsized role in TB endemic settings. Modeling studies

[1]Centre for Respiratory Diseases Research, Kenya Medical Research Institute, Nairobi, Kenya. [2]Department of Global Health, University of Washington, Seattle, WA, USA. [3]Department of Medicine, University of Washington, Seattle, WA, USA. [4]Division of Intramural Research, National Heart, Lung and Blood Institute (NHLBI), National Institutes of Health, Bethesda, MD, USA. e-mail: dhorne@uw.edu

suggest greatly amplified returns by focusing control efforts on the minority of persons with TB who are most infectious[12]. Although the existence of "superspreaders" in Mtb transmission remains unproven, investigations into this phenotype may provide insights into TB pathophysiology and inform the development, implementation, and evaluation of targeted public health interventions to eliminate TB[3].

Epidemiologic studies of Mtb infectiousness suggest that several host characteristics, including younger age, higher bacillary load, cough features, and greater contact time, are associated with increased transmission[13–18]. Overall, these associations are weak and suggest that additional factors regulate infectivity. Studies by Wells and Riley from the pre-chemotherapeutic era demonstrated that a subset of patients with TB generated small droplets capable of infecting guinea pigs[15,19–21]. More recent studies focused on people living with HIV (PLWH) similarly found that a subset of patients transmitted Mtb[22]. Compared to sputum smear and culture assessment, direct measurement of Mtb aerosolization with a cough aerosol sampling system (CASS)[23–28] is a more accurate predictor of transmission among household contacts (HHCs)[24,26]. A recent study suggested that sputum bacillary load and clinical characteristics, including fewer symptoms and a stronger cough, predict Mtb aerosolization[29]. Studies using a Respiratory Aerosol Sampling Chamber (RASC), which directly measures viable Mtb in bioaerosols[30–32], as well as those that used face mask sampling[33] suggested that tidal breathing and non-cough mechanisms may be important drivers of transmission[34]. Although these studies demonstrate features of airborne Mtb transmission, the biological mechanisms remain poorly understood, including whether immune pathways modulate aerosolization. In addition, while CASS and RASC are important research tools, a simple diagnostic test or algorithm that identifies superspreaders has not been developed and would have a significant impact on efficient resource allocation for TB control[35,36].

To address these gaps in knowledge, we designed the TB Aerobiology, Immunology and Transmission (TBAIT) study. We performed a prospective, longitudinal study using CASS, enrolled 142 participants with pulmonary TB, and assessed the association of clinical, microbiologic, and immunologic characteristics with Mtb aerosolization. In addition, we measured interferon-gamma release assay (IGRA) responses in HHCs to assess Mtb transmission risk by cough aerosol culture (CAC) status of the index participant.

Results

Cough aerosol culture results among participants with pulmonary TB

To examine the biology of Mtb aerosolization and transmission, we enrolled 142 individuals with microbiologically confirmed pulmonary TB in Nairobi, Kenya. The median age of participants was 35 years (interquartile range (IQR), 27–44) and ranged from 18 to 100 years; 27% were women (Table 1). All participants identified as Black. Most participants were symptomatic including cough in 95% and weight loss in 85%. Most participants had medium (32%) or high (39%) GeneXpert semi-quantitative grades and findings of cavitations on chest x-rays (70%). Only 11% of participants were living with HIV, of whom one-half were taking antiretroviral therapy at enrollment. As compared to participants identified through passive case finding ($n = 116$), those identified through active case finding ($n = 26$) were older (mean 43 vs. 35 years, p-value 0.005), less likely to report a cough (81% vs. 98%, p-value < 0.001), less likely to have cavitary disease on chest x-ray (38% vs 77%, p-value < 0.001) and had higher mean GeneXpert Ct values (24.0 vs. 19.2, p-value < 0.001); there was no difference in the frequency of women or PLWH.

The median CRP value among all participants was high at 67 mg/L (IQR 26–96). Forty-three participants (30%) were cough aerosol culture-positive and had higher measures of bacillary burden, including higher AFB-smear grades and shorter time to detection of Mtb growth in liquid culture, compared to cough aerosol culture-negative participants (Table 1). Cough aerosol culture-positive participants were more likely to have cavitary lung disease and at least two quadrants affected by TB on chest x-ray than cough aerosol culture-negative persons. GeneXpert cycle threshold was significantly lower and semi-quantitative grade significantly higher among cough aerosol culture-positive compared to cough aerosol culture-negative participants. More cough aerosol culture-positive participants (91%) had GeneXpert semi-quantitative grades of medium or high compared to 62% of cough aerosol culture-negative individuals. The correlation between GeneXpert cycle threshold and cough aerosol culture status was −0.44, a moderate correlation which suggests that additional factors determine aerosolization.

Index cough aerosol culture positivity associated with IGRA positive result in household contacts

We next evaluated IGRA results in household contacts to determine whether cough aerosol culture status is associated with infectiousness. Of the 142 participants with TB, 30 lived alone and were not eligible for the household contact study, and 64 had household members who could not be contacted, or the household members declined enrollment. Index participants ($N = 48$) with enrolled household members compared to index participants ($N = 94$) without enrolled household members did not differ by age (p-value 0.12), HIV status (p-value 0.74), MUAC (p-value 0.60), or CASS status (0.17). (Supplementary Table 5) However, there were differences in the proportion of women (42% vs. 20%, p-value 0.007), Xpert Ct values (20.5 vs. 18.1, p-value 0.01), and the frequency of cavitations on chest X-rays (58% vs. 76%, p-value 0.04) between those with and without enrolled household members, respectively. We enrolled 129 household contacts and obtained QFT tests in 116 (13 declined phlebotomy or had unsuccessful phlebotomy) (Table 2). In bivariate analyses, household contacts did not differ in gender or age by cough aerosol culture status of the index participant. The overall frequency of QFT-positive results among household contacts was 60% and differed by cough aerosol culture status of the index participant: excluding indeterminate results, 85%, and 53%, for contacts of cough aerosol culture-positive and -negative participants, respectively (p-value 0.006). Among household contacts less than 10 years of age, QFT-positive results were also more common (p-value 0.01) among contacts of CAC-positive persons (9 of 10 participants, 90%) than contacts of CAC-negative persons (14 of 31 participants, 45%). Cough aerosol culture-positive contacts had a higher mean IGRA IFNγ level compared to cough aerosol culture-negative individuals (p-value 0.001, 4.25 IU/ml vs. 1.35 IU/ml). Among participants who were QFT-positive, the mean IFN-γ levels were significantly higher among contacts of cough aerosol culture-positive persons ($n = 22$, 4.76, SD 2.58) compared to contacts of cough aerosol culture-negative persons ($n = 48$, 3.36, SD 1.97, $P = 0.02$).

To establish the background prevalence of QFT-positive results in Nairobi residents, we evaluated QFT status in adult outpatients who presented for health care due to cough ($n = 45$, median age 39 years (IQR, 34-45)) and to HIV prevention and care centers ($n = 121$ PLWH; $n = 122$ without HIV; overall median age 36 years (IQR, 28-44). After excluding indeterminate results ($n = 2$ and $n = 34$, respectively), we found that the prevalence of QFT-positive results was, 51% ($n = 22$) and 51% ($n = 121$), respectively. Among the 45 persons who presented to health care with a cough, QFT status did not differ by HIV status (PLWH = 17, p-value 0.66). In participants recruited at HIV care centers, QFT results differed by HIV status: 43% of PLWH (48 of 111) and 60% of persons without HIV (59 of 98) (p-value 0.01) were QFT positive. The estimated Nairobi background prevalence of QFT test positivity did not differ from contacts of cough aerosol culture-negative participants (p-value 0.33) but was significantly lower than contacts of cough aerosol culture-positive participants (p-value < 0.001).

https://doi.org/10.1038/s41467-024-52122-x

Table 1 | Characteristics of Participants with TB (Index participants) overall and by cough aerosol culture (CAC) status

Characteristics	Total *n* = 142 N(%) or (med, IQR)	CAC + *n* = 43	CAC- *n* = 99	Odds Ratio	*p*-value
Age, years	35 (27, 44)	32 (24, 39)	36 (28, 48)	0.95 (0.91 – 0.98)	0.006
Female	39 (27%)	8 (19%)	31 (31%)	0.50 (0.21 – 1.21)	0.12
BMI (*n* = 141)	19 (17, 21)	19 (18, 21)	19 (17, 21)	1.02 (0.92 – 1.14)	0.69
MUAC[a], cm (*n* = 141)	23 (21, 25)	24 (21, 26)	22 (21, 25)	1.14 (1.00 – 1.30)	0.04
PLWH[b]	16 (11%)	4 (9%)	12 (12%)	0.74 (0.23 – 2.45)	0.88
Taking ART?	8 (50%)	4 (100%)	4 (33%)		0.02
CD4 T-cell count, cells/μL (*n* = 15)	193 (67, 231)	218 (105, 397)	178 (67, 216)	1.00 (1.00 – 1.01)	0.63
Prior history of TB (*n* = 140)	27 (19%)	8 (19%)	19 (20%)	0.94 (0.37 – 2.35)	0.89
Enrolled through active case finding	26 (18%)	0	26 (26%)	–	–
Household residents	2 (1, 4)	2 (1, 4)	2 (1, 4)	–	0.94
Chest X-ray with cavitary findings	99 (70%)	38 (88%)	61 (62%)	4.73 (1.71 – 13.08)	0.003
Number of chest X-ray quadrants with TB involvement (*n* = 140)	–	–	–	1.50 (1.02 – 2.20)	0.04
0	1 (1%)	0	1 (1%)	–	–
1	47 (34%)	8 (19%)	39 (40%)	–	–
2	53 (38%)	18 (42%)	35 (36%)	–	–
3	28 (20%)	15 (35%)	13 (13%)	–	–
4	11 (8%)	2 (5%)	9 (9%)	–	–
Cough (*n* = 141)	133 (95%)	43 (100%)	91 (93%)	–	0.07
Fever (*n* = 141)	92 (65%)	35 (81%)	57 (58%)	3.15 (1.32 – 7.49)	0.10
Weight loss (*n* = 141)	120 (85%)	41 (95%)	79 (81%)	1.00 (0.97 – 1.03)	0.76
Night sweats	108 (77%)	36 (84%)	72 (73%)	1.86 (0.74 – 4.69)	0.19
Cough duration (days)(*n* = 134)	8 (4,12)	8 (4, 12)	6 (4, 12)	1.03 (1.0 – 1.07)	0.10
LCQ Score[c] (*n* = 140)	14.5 (12.2, 16.7)	13.1 (10.4, 14.4)	15.5 (12.8, 17.2)	0.81 (0.71 – 0.92)	0.001
CPF[d], mL	260 (170, 370)	225 (170, 350)	270 (170, 370)	1.00 (1.0 – 1.0)	0.23
Current tobacco use	32 (23%)	9 (21%)	23 (23%)	0.86 (0.36-2.06)	0.74
Xpert Ct	18.8 (16.6, 23.3)	16.4 (14.6, 18.1)	19.7 (17.8, 25.1)	0.73 (0.63 – 0.84)	0.000006
Xpert Semiquantitative Grade	–	–	–	2.12 (1.35 – 3.22)	0.001
Trace	4 (3%)	0	4 (4%)	–	–
Very Low	9 (6%)	1 (2%)	8 (8%)	–	–
Low	28 (20%)	3 (7%)	25 (26%)	–	–
Medium	45 (32%)	15 (35%)	30 (30%)	–	–
High	56 (39%)	24 (56%)	32 (32%)	–	–
Rifampin resistance detected	1 (1%)	0	1 (1%)	–	–
Sputum Smear Grade	–	–	–	1.62 (1.16 – 2.26)	0.004
Negative	15 (11%)	1 (2%)	14 (14%)	–	–
Scanty	9 (6%)	2 (5%)	7 (7%)	–	–
1+	33 (23%)	5 (12%)	28 (28%)	–	–
2+	35 (25%)	16 (37%)	19 (19%)	–	–
3+	50 (35%)	19 (44%)	31 (31%)	–	–
TTD[e] (days) (*n* = 136)	5 (4, 9)	4 (3, 5)	6 (4, 10)	0.75 (0.63 – 0.88)	0.0005
Sputum Appearance	–	–	–	–	–
Thin	46 (32%)	10 (23%)	36 (36%)	REF	
Thick	91 (64%)	32 (74%)	59 (60%)	1.95 (0.86 – 4.44)	0.11
Bloody	5 (4%)	1 (2%)	4 (4%)	0.90 (0.09 – 8.98)	0.93
CRP, mg/dL (*n* = 141)	67 (26, 96)	88 (57, 131)	51 (15, 86)	1.02 (1.01 – 1.03)	0.00005
WBC	7.1 (5.7, 9.1)	7.7 (6.4, 10.1)	6.8 (5.6, 8.6)	1.19 (1.03 – 1.36)	0.02
Granulocytes	5.1 (3.8, 7.0)	5.7 (4.7, 7.8)	4.8 (3.6, 6.6)	1.20 (1.03 – 1.39)	0.02
Lymphocytes	1.4 (1.1, 1.8)	1.5 (1.1, 1.9)	1.4 (1.0, 1.8)	1.20 (0.72 – 1.99)	0.49
Monocytes	0.4 (0.2, 0.5)	0.3 (0.3, 0.5)	0.4 (0.2, 0.5)	0.77 (0.29 – 2.07)	0.61
HbA1C, % (*n* = 141)	5.7 (5.3, 6.0)	5.8 (5.5, 6.1)	5.6 (5.1, 6.0)	1.1 (0.79 – 1.45)	0.67
Hgb	13 (11.9, 14.2)	13.0 (12.0, 13.9)	13.1 (11.8, 14.3)	1.00 (0.84 – 1.19)	1

[a]MUAC, Mid-upper arm circumference; [b]PLWH, a person living with HIV; [c] LCQ, Leicester Cough Questionnaire; [d]CPF, Cough peak flow; [e]TTD, Time to detection of Mtb growth in liquid media
Characteristics were compared by cough aerosol culture status using bivariate logistic regression, except for cough, which was tested using a chi-square statistic. No adjustments were made for multiple comparisons. Source data are provided as a Source Data file.

https://doi.org/10.1038/s41467-024-52122-x

Table 2 | Household Contact Characteristics overall and by index cough aerosol culture (CAC) status

Characteristics	Total *n* = 129 N(%) or (med, IQR)	CAC + *n* = 27	CAC- *n* = 102	*p*-value
Household Contact Characteristics (*n* = 129)				
Age, years (med, IQR)	12 (5, 29)	12 (5, 28)	13 (5, 32)	0.54
Age groups				
< 5 years	31 (24%)	6 (22%)	25 (25%)	0.58
5–10 years	20 (15%)	6 (22%)	14 (14%)	–
10–15 years	19 (15%)	5 (19%)	14 (14%)	–
> 15 years	59 (46%)	10 (37%)	49 (48%)	–
Women	66 (51%)	15 (56%)	51 (50%)	0.61
BMI[b] (*n* = 122)	19 (16, 24)	19 (15, 25)	20 (16, 24)	0.39
MUAC[c], cm (*n* = 123)	21.5 (16, 25.5)	19.2 (15.5, 29.5)	22 (16.2, 25)	0.87
PLWH[a] (*n* = 99)	4 (4%)	1 (5%)	3 (4%)	0.76
Characteristics of Index Participants (*n* = 48) with whom Household Contacts Resided				
Age	34 (27, 47)	37 (29, 41)	33 (27,47)	0.49
Women	48 (37%)	10 (37%)	38 (37%)	0.98
PLWH[a]	17 (13%)	2 (7%)	15 (15%)	0.32
BMI[b] (*n* = 126)	19 (18, 23)	18 (18, 21)	19 (18, 23)	0.21
MUAC[b] (*n* = 126)	23.2 (20.8, 25.6)	22.9 (22.4, 25.5)	23.3 (20.8, 25.8)	0.38
Xpert Ct (*n* = 129)	20.5 (17.6, 26.7)	19.0 (16.6, 20.4)	22.7 (17.8, 27.3)	0.0002
Xpert Grade (*n* = 129)	–	–	–	0.003
Trace	3 (2%)	0	3 (3%)	–
Very Low	17 (13%)	0	17 (17%)	–
Low	30 (23%)	2 (7%)	28 (27%)	–
Medium	53 (41%)	19 (70%)	34 (33%)	–
High	26 (20%)	6 (22%)	20 (20%)	–
Smear Grade (*n* = 129)	–	–	–	0.25
Negative	23 (18%)	4 (15%)	19 (19%)	–
Scanty	5 (4%)	1 (4%)	4 (4%)	–
1+	42 (33%)	7 (26%)	35 (34%)	–
2+	27 (21%)	10 (37%)	17 (17%)	–
3+	32 (25%)	5 (19%)	26 (26%)	–
TTD[d] (days) (*n* = 122)	5.5 (4, 12)	5 (4, 6)	6 (4, 12)	0.04
Cavitary CXR (*n* = 126)	74 (59%)	24 (89%)	50 (51%)	0.001
CRP (*n* = 124)	46 (19, 75)	51 (26, 96)	46 (7, 75)	0.11
CXR quadrants (*n* = 126)	–	–	–	0.001
0	0	0	0	–
1	49 (39%)	6 (22%)	43 (43%)	–
2	54 (43%)	11 (41%)	43 (43%)	–
3	18 (14%)	10 (37%)	8 (8%)	–
4	5 (4%)	0	5 (5%)	–
Number of contacts for each index participant (% of households)	–	–	–	0.02
1 HHC	16 (33%)	4 (36%)	12 (32%)	–
2 HHCs	9 (19%)	3 (27%)	6 (16%)	–
3 HHCs	8 (17%)	1 (9%)	7 (19%)	–
4 HHCs	7 (15%)	1 (9%)	6 (16%)	–
5 HHCs	5 (10%)	2 (18%)	3 (8%)	–
6 HHCs	3 (6%)	0	3 (8%)	–
QuantiFERON (QFT) Result in HHC (*n* = 116)				
QFT-neg	44 (38%)	3 (12%)	41 (46%)	0.006
QFT-pos	70 (60%)	22 (85%)	48 (53%)	–
Indeterminate	2 (2%)	1 (4%)	1 (1%)	–
Mean IFN-γ (IU/mL) (SD) (indeterminate excluded)	2.34 (2.55)	4.18 (2.89)	1.83 (2.20)	0.0002
Median IFN-γ (IU/mL) (IQR) (indeterminate excluded)	1.02 (0.02 – 4.45)	4.28 (1.77 – 5.91)	0.71 (0 – 3.56)	–
QFT-positive HHCs, Mean IFN-γ (IU/mL) (SD) (*n* = 70)	3.80 (2.25)	4.76 (2.58)	3.36 (1.97)	0.02
QFT-positive HHCs, Median IFN-γ (IU/mL) (IQR) (*n* = 70)	3.71 (1.85-5.06)	4.59 (3.35–6.97)	3.46 (1.82–4.79)	–

[a]PLWH, person living with HIV; [b]BMI, Body mass index; [c]MUAC, Mid-Upper Arm Circumference; [d]TTD = Time to detection of Mtb growth in liquid media

Using bivariate logistic regression, characteristics were compared by cough aerosol culture status. No adjustments were made for multiple comparisons. The source data are provided as a Source Data file.

Using mixed effects logistic regression models, we next evaluated bivariate associations between measures of index case TB infectiousness and QFT results in contacts. (Table 3) Using multivariable models with random intercepts to account for clustering by index participant, we evaluated predictors of interest for associations with QFT positivity in household contacts and found that the best-performing model included index cough aerosol culture status and HIV status, and household contact's age. (Table 3).

Models of index characteristics associated with cough aerosol culture status

We evaluated host characteristics associated with cough aerosol culture status based on a conceptual framework that included potential confounders (Table 1 and Supplementary Fig. 2). On bivariate analyses, younger age, higher MUAC, cavitary disease on chest X-ray, TB-related abnormalities in more chest X-ray quadrants, and higher cough-related impairment effects on quality of life (lower Leicester Cough Questionnaire score) were associated with cough aerosol culture-positive status. Factors associated with higher bacillary burden, including lower GeneXpert Ct value, higher GeneXpert grade, higher sputum smear grade, and shorter time-to-detection of Mtb culture growth, were all associated with cough aerosol culture-positive status. Higher WBC counts (total and granulocyte count) and CRP levels were also associated with cough aerosol culture-positive status (Fig. 1a). We compared multivariable models for predictors associated with cough aerosol culture-positive status and found that the best-performing model included lower GeneXpert Ct value, lower age, higher CRP level, higher MUAC, and shorter TTD of culture growth (Table 4). The coefficient of determination (R^2) of this model was 0.37, suggesting that other factors, in addition to the evaluated independent variables, determine TB infectiousness.

Performance of clinical prediction rule to identify cough aerosol culture-positive persons

We developed and evaluated the performance of a clinical prediction rule to predict cough aerosol culture-positive persons as a means to

Table 3 | Mixed methods bivariate & multivariable analyses of index and household contact characteristics associated with QFT result in household contacts (excluding indeterminate results)

Predictors[a]	OR	95% CI	*p*-value	aOR	95% CI	*p*-value
Household Contact Characteristics						
Age, years	1.04	1.00–1.07	0.049	1.03	0.99–1.06	0.10
Women	1.03	0.41–2.58	0.95	–	–	–
PLWH[a]	2.16	0.16–28.45	0.56	–	–	–
Index Characteristics						
Cough aerosol culture-positive	8.51	1.62–44.78	0.01	8.81	1.46–53.20	0.02
Age, years	1.02	0.98–1.06	0.28	–	–	
Women	1.56	0.47–5.15	0.46	–	–	
PLWH[a]	0.09	0.01–0.65	0.02	0.12	0.02–0.88	0.04
Xpert Ct	0.89	0.78–1.00	0.06	–	–	–
Xpert semi-quantitative grade	1.50	0.85–2.63	0.16	–	–	–
TTD[b] (days)	0.88	0.73–1.06	0.18	–	–	–
AFB-smear grade	1.40	0.93–2.12	0.11	–	–	–
CRP level, mg/dL	1.01	1.00–1.03	0.10	–	–	–
Cavitary chest X-ray	2.85	0.93–8.69	0.07	–	–	–
CXR Quadrants	1. 38	0.69–2.75	0.36	–	–	–
Body Mass Index	0.98	0.85–1.13	0.79	–	–	–
MUAC[†]	0.94	0.79–1.12	0.51	–	–	–

[a]PLWH, person living with HIV; [b]TTD, Time to detection of Mtb growth in liquid media; [†]MUAC, Mid-upper arm circumference

[a]Predictors were selected based on our conceptual model (see Supplementary Fig. 1), tested in forward selection stepwise multivariable models with random intercepts to account for clustering by index participant, and compared with likelihood ratio test. The best-performing model with adjusted odds ratios is shown in the final three columns. No adjustments were made for multiple comparisons. Source data are provided as a Source Data file.

identify those who are likely to be highly infectious. We included predictors from our multivariable logistic regression model that are readily available to a clinician during the initial visit. For this reason, we removed the time-to-detection of culture growth but kept the CRP level as there are point-of-care versions of this test[37]. We also used Xpert Ultra semi-quantitative grade, which is the result generally reported to clinicians and is derived from the Ct value. Using the coefficients obtained in the multivariable logistic regression analysis, we derived the following prediction equation for cough aerosol culture-positive status where outcomes were coded according to weighted scores based on the beta coefficient: CRP > 42.9 + Age < 43.8 years + MUAC > 22.8 cm + Xpert semi-quantitative grade (high, medium, or low/very low/trace). A total risk score is calculated by adding the risk points and has an optimum cut point of 15 points for predicting cough aerosol culture-positive status with an estimated sensitivity of 0.86 (95% CI, 0.74–0.95) and specificity of 0.65 (95% CI, 0.56–0.74). (Supplementary Tables 2 and 3) The prediction rule based on the calculated risk score had an area under the receiver operating curve (AUROC) of 0.84 (95% CI: 0.77–0.91) (Supplementary Fig. 3). The model's Somers' D_{xy} index was 0.71; the equivalent in bootstrap validation was 0.66, with an optimism estimate of 0.051, indicating good stability of the model in internal validation.

Sputum and plasma cytokine analysis of associations with cough aerosol culture positivity, bacterial load, and cavitary lung disease

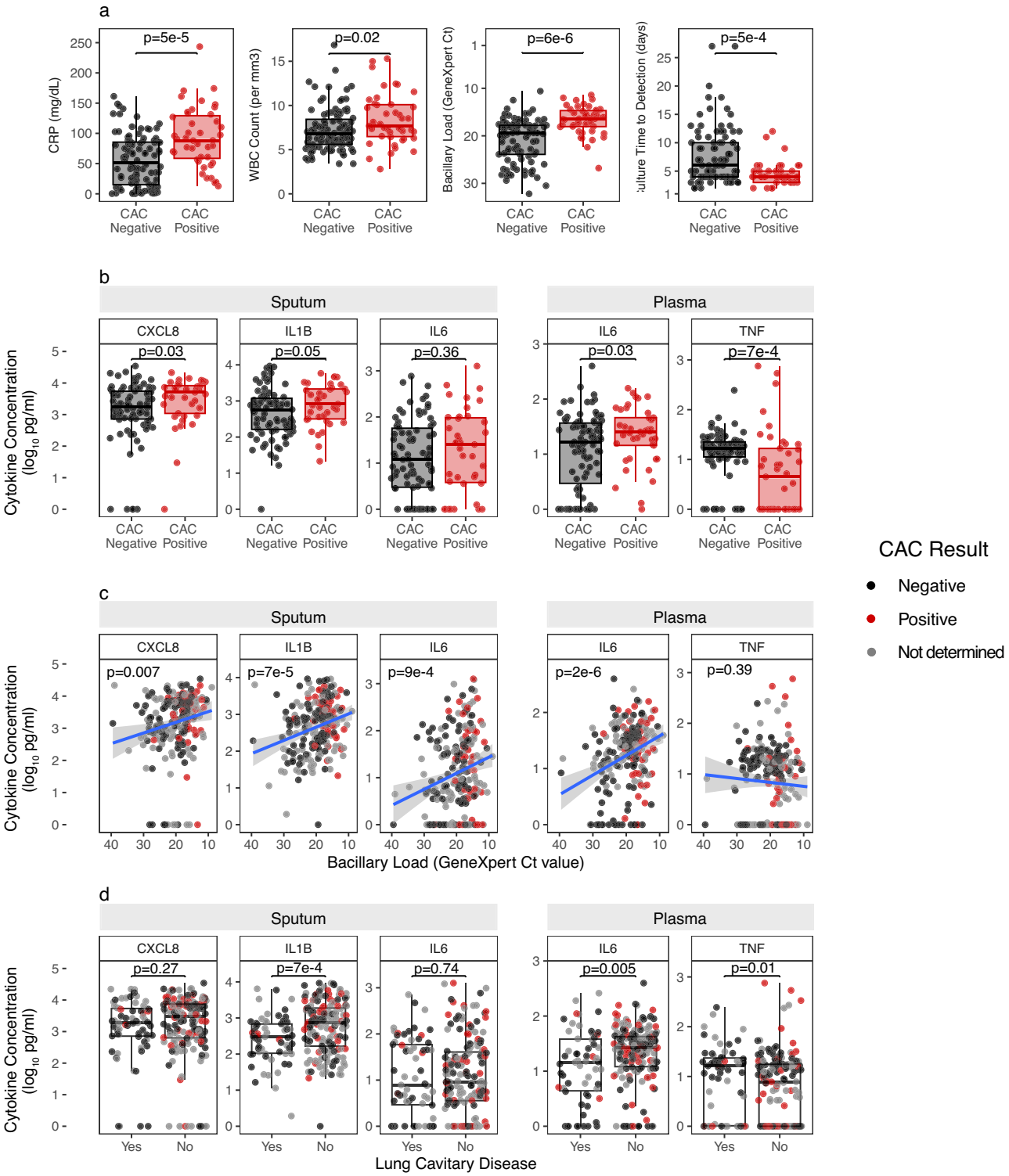
To further examine the association of inflammatory markers with cough aerosol culture-positive status, we analyzed sputum and plasma levels of CXCL8 (IL-8), IL-1β, TNF, and IL-6 (Fig. 1b and Supplementary Table 4). TNF in sputum and IL-1β and CXCL8 in plasma were nearly undetectable in preliminary subgroup testing and were not examined further. The sputum concentration of CXCL8 was significantly higher (*p*-value 0.03) among the cough aerosol culture-positive compared to cough aerosol culture-negative participants. Sputum IL-1β (*p*-value 0.052) and IL-6 (*p*-value 0.36) were not different by CAC status (Fig. 1b). Higher IL-6, CXCL8, and IL-1β levels were associated with greater bacillary burden (measured by GeneXpert cycle threshold, *p*-value 0.0009, 0.007 and 0.00007, respectively, Fig. 1c). IL-1β, but not IL-6 or CXCL8, was positively associated with cavitary lung disease (*p*-value 0.0007) across cough aerosol culture status.

Plasma IL-6 was associated with bacillary burden ($p = 1.69 \times 10{-}6$) and cavitary disease ($p = 0.005$) and was higher in CAC-positive compared to CAC-negative participants ($p = 0.03$). Plasma TNF was associated with cavitary disease ($p = 0.013$), but not bacillary burden, and was lower in CAC-positive compared to CAC-negative participants (*p*-value 0.0007). For cytokines that had a significant association with cough aerosol status in bivariate models (sputum CXCL8, plasma TNF, plasma IL-6), we evaluated associations in multivariate logistic regression models in which we adjusted for age, sex, cavitary disease on chest x-ray and GeneXpert Ct value. We found that plasma TNF was independently associated with CAC status (adjusted *p*-value 0.007, Supplementary Table 4).

Whole blood transcriptomic signatures and analysis of associations with cough aerosol culture positivity and sputum bacterial load

We selected 58 subjects for whole blood transcriptomic analysis which were a subgroup of the full cohort of 142. We initiated this experiment after 29 CAC + and 29 CAC- subjects were available for analysis from the initial phase of enrollment. We found no differences in clinical and biologic variables (age, sex, HIV status, Xpert Ct value, cavitary CXR, and degree of lung involvement (CXR quadrants)) when comparing participants included in the transcriptomic subgroups (CAC + and CAC-) with the remaining cohort (Supplementary Tables 6 and 7). We used a genome-wide approach to determine whether a specific host inflammatory signature was associated with cough aerosol culture-positive status independent of other diseases. We measured whole blood RNA-seq profiles from a pre-treatment sample of persons with pulmonary TB and found differentially expressed genes associated with cough aerosol culture-positive status (100 genes with FDR < 0.2).

https://doi.org/10.1038/s41467-024-52122-x



While a strong majority of the genes (78%) were best fit by models including cough aerosol culture status alone, 16% of genes were best fit with the inclusion of bacillary load with a smaller percentage best explained by more complex models with age and cavitary disease (Supplementary Fig. 4a). After a stepwise evaluation of covariates and assessment of model fit (based on improvement of >10% of genes), bacillary load (measured by Ct value) was the only covariate included in the final model. We examined covariate-adjusted expression values and found that no differentially expressed genes (DEGs) were independently associated with cough aerosol culture positivity (Fig. 2a, b and Supplementary Fig. 2c). In contrast, many DEGs remained associated with bacterial burden (1129 genes with FDR < 0.2; 40 genes with FDR < 0.05; min FDR = 0.0008, Supplementary Table 7). Based on these findings, DEGs are associated with several features of TB presentation but are not independently associated with cough aerosol culture-positive status after adjusting for bacillary burden.

We next used Gene Set Enrichment Analysis (GSEA) to examine whether transcriptional signatures were associated with cough aerosol culture-positive status. Using an adjusted model with a bacillary burden as a covariate and assessing for concordant directionality of effect,

Fig. 1 | Association of blood and sputum inflammatory and microbiologic markers with cough aerosol culture positivity. Inflammatory and microbiologic markers were measured in blood and sputum at baseline visits and correlated with cough aerosol culture positivity. **a** Serum CRP, Blood WBC, sputum bacillary load determined by GeneXpert Ct value, and sputum culture time to detection ($n = 99$ CAC-; $n = 43$ CAC +). **b**–**d** Sputum and plasma cytokines measured by ELISA; CXCL8, IL1B, IL6, and TNF measurements at diagnosis were correlated with cough aerosol culture positivity (**b**), bacillary (**c**), and chest x-ray lung cavitation load (panels b-d: sputum CXCL8: $n = 81$ CAC-, $n = 37$ CAC +; sputum IL1B: $n = 80$ CAC-, n = 37 CAC +; sputum IL6: $n = 82$ CAC-, $n = 38$ CAC +; plasma IL6 & TNF: $n = 91$ CAC-, $n = 41$ CAC +) (**d**). For **c**, **d** cough aerosol culture-positive and cough aerosol culture-negative individuals are depicted with red or black circles, respectively. Figures (**a**), (**b**), and (**d**) are Tukey-style box plot statistics with the median line, hinges indicating the 1st and 3rd quartiles, and whiskers extending to maximum and minimum values within 1.5x the interquartile range of the upper and lower hinges. The graphs in (**c**) display the linear model fit to the data with shaded error regions indicating the 95% confidence interval of the fit. Analyses shown in 1(**a**) panels reflect bivariate logistic regression described in Table 1, those shown in 1(**b**) and 1(**d**) panels used two-sided Wilcoxon rank sum test with continuity correction, and 1(**c**) panels reflect linear models fit to log10 transformed cytokine concentration. ELISA was performed with technical duplicates on the same sample with highly discordant samples removed from further analysis; results were averaged for each participant. Gray circle, unknown cough aerosol culture status. Source data are provided as a Source Data file. $^{*}p < 0.05$, $^{**}p < 0.01$, $^{***}p < 0.001$.

several gene sets were associated with bacillary load or cough aerosol culture-positive status and represented a diverse set of cellular processes (Fig. 2c). Although there was overlap among cough aerosol culture-positive and bacillary burden associations, the majority of gene sets were associated with concordant directionality with either, but not both, traits (Fig. 2c). We found eight Hallmark gene sets that were associated either exclusively with cough aerosol culture-positive status, or with opposite directionality as bacillary burden. Hallmark_Angiogenesis was associated with cough aerosol culture-positive status, while seven gene sets were associated with cough aerosol culture-negative status (Fig. 2c and Supplementary Fig. 5). There were 10 leading-edge genes in the Angiogenesis gene set with functional associations with cellular proliferation, ligand:receptor interactions, and the extracellular matrix. (Fig. 2d and Supplementary Table 8).

Table 4 | Multivariable model of index characteristics associated with cough aerosol culture positivity

Characteristics	OR	95% CI	p-value	aOR	95% CI	p-value
MUAC,[a] cm	1.14	1.00–1.30	0.04	1.42	1.17–1.73	0.0004
CRP level, mg/dL	1.02	1.01–1.03	< 0.001	1.02	1.01–1.03	0.001
Xpert Ct	0.73	0.63–0.84	< 0.001	0.76	0.64–0.90	0.002
Age, years	0.95	0.92–0.99	0.006	0.94	0.90–0.99	0.02
TTD[b], days	0.75	0.63–0.88	< 0.001	0.80	0.64–1.00	0.05
Female	0.50	0.21–1.21	0.12	–	–	–
BMI	1.02	0.92–1.14	0.69	–	–	–
Sputum smear grade	1.62	1.16–2.26	0.005	–	–	–
Xpert semi-quantitative grade	2.12	1.35–3.22	0.001	–	–	–
PLWH[c]	0.74	0.23–2.45	0.88	–	–	–
Fever	3.15	1.32–7.49	0.10	–	–	–
Cough Duration	1.03	1.00–1.07	0.10	–	–	–
LCQ[d] Score	0.81	0.71–0.92	0.001	–	–	–
CPF[e]	1.00	1.00–1.00	0.23	–	–	–
Cavitary Chest X-ray	4.73	1.71–13.08	0.003	–	–	–
Chest x-ray quadrants	1.50	1.02–2.20	0.04	–	–	–
WBC	1.19	1.03–1.36	0.02	–	–	–

[a]MUAC, Mid-Upper Arm Circumference; bTTD, Time to detection of Mtb growth in liquid media; [c]PLWH, Person living with HIV; [d]LCQ, Leicester Cough Questionnaire; [e]CPF, Cough peak flow
Predictors were selected based on our conceptual model (see Supplementary Fig. 1), tested in forward selection stepwise multivariable models, and compared with the likelihood ratio test. The best-performing model with adjusted odds ratios is shown in the final three columns. No adjustments were made for multiple comparisons. Source data are provided as a Source Data file.

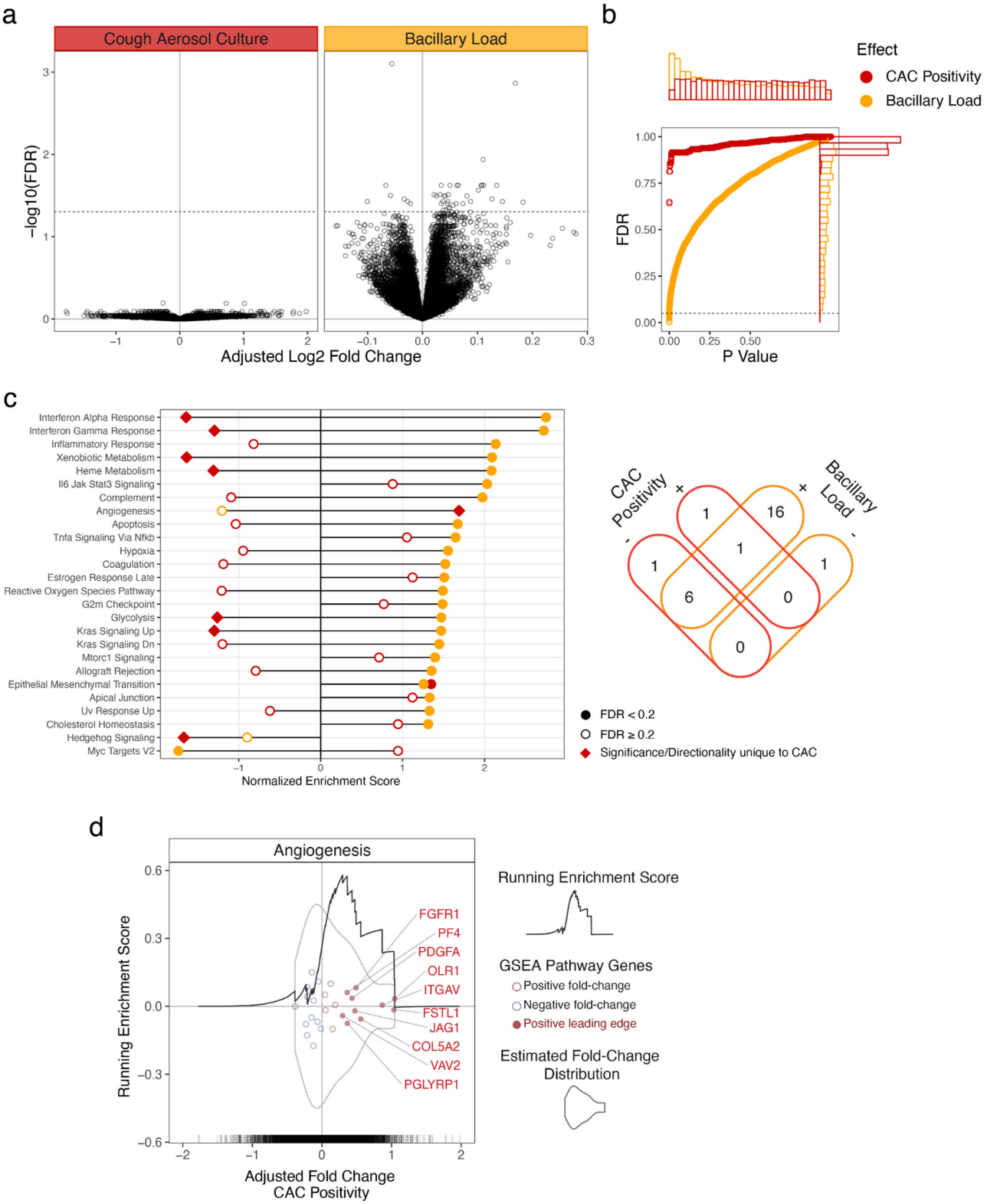
Discussion

The primary findings of TBAIT suggest that host inflammatory signatures are associated with Mtb aerosolization independent of bacillary load and cavitary lung disease. We extend previous studies that determined several non-immunologic host factors that are associated with cough aerosol culture status[25,29]. We found that higher serum CRP levels, sputum and plasma cytokines, and whole blood transcriptional signatures were associated with Mtb aerosolization. These data highlight potential insights into the biology of Mtb transmission events as well as biomarkers to identify highly infectious individuals. Based on our findings and prior studies, cough aerosol cultures are superior to sputum smear analysis in predicting Mtb transmission events and are likely the best estimators of TB infectiousness currently[24,26–28]. Although cough aerosol culture-negative patients may transmit Mtb, cough aerosol cultures allow for a more accurate assessment of relative infectiousness than traditional measures of sputum bacillary load. Prior studies found that bacillary burden, mucoid sputum, stronger cough, and higher Karnofsky performance score were associated with cough aerosol culture positivity[25,29].

Theron et al. in the largest cough aerosol culture study of TB patients to date[29], found that higher peak cough flow rate, higher bacillary load, lack of HIV infection, and lower "TB symptom score" were independent risk factors for cough aerosol culture positivity. In the Theron study, a lower TB symptom score indicated a lower burden of findings attributable to TB disease as it is a summation of points for TB-related symptoms (cough, hemoptysis, dyspnea, chest pain, fever, night sweats), TB-related signs (anemia, tachycardia), lung auscultation findings, and malnutrition (low BMI, low MUAC). Our study adds the association between cough aerosol culture-positive status and higher serum CRP levels, an acute phase reactant that is produced in the liver and is a non-specific marker of inflammation. CRP has been extensively evaluated as a triage test for TB, particularly among PLWH[38], is endorsed by WHO as a TB screen, and is available as a point-of-care test[37]. We also identified an independent association between lower plasma TNF levels and culture-positive cough aerosols. Our finding of an association between cough aerosol status and MUAC, but not BMI, may seem counterintuitive and may represent a statistical artifact. Alternatively, we hypothesize that MUAC and BMI measure related but different aspects of nutritional status: BMI reflects fat mass while MUAC measures fat and muscle mass.[39]. Low MUAC was a better predictor than BMI of all-cause mortality in older adults[39–41], and in persons with TB, MUAC (unlike BMI) was independently associated with cavitary lung disease[42], Cachexia, severe weight loss that results from muscle atrophy and the loss of adipose tissue, may correlate with differential immune cell responses[43]. For example, using parasitic infections in the murine model, mice deficient for CD4 + T cells experienced muscle and fat wasting as opposed to CD8 + T cell-deficient mice, which experienced only fat wasting[44].

Taken together, these findings suggest common features of a cough aerosol culture-positive phenotype: younger persons with few symptoms (aside from cough), preserved muscle mass, a higher

https://doi.org/10.1038/s41467-024-52122-x



bacillary burden, and higher systemic inflammation. If accurate, the proposed phenotype of highly infectious persons with TB would likely describe individuals who are younger[16–18,29], and active rather than moribund[25,29], increasing the opportunities for TB transmission events. We developed a clinical prediction tool to identify highly infectious persons with TB (cough aerosol culture positive) that could have utility in TB control interventions, for example, to inform contact investigations and TB preventive therapy administration, and to improve precision in clinical trials of TB preventive therapy and vaccines. Like other clinical prediction rules, the model performance is predicated on having data for all of the characteristics that contribute to the risk score. While promising, the risk score that we present requires external validation and should not be used at this time for clinical (non-research) purposes.

Potential mechanisms underlying increased Mtb aerosolization and transmission include high sputum bacillary loads, Mtb strain

Fig. 2 | Whole blood transcriptional signatures and association with cough aerosol culture positivity and bacillary burden. a Differential gene expression associated with cough aerosol culture positivity and bacillary load. The volcano plot shows the gene-wise covariate-adjusted log2 fold change associated with each effect as well as each gene's Benjamini-Hochberg adjusted false discovery rate (FDR). **b** Distributions of unadjusted p-values (*x*-axis) and FDR (*y*-axis) for each effect. **c** Gene set enrichment of Hallmark pathways in differential gene expression of cough aerosol culture positivity (red) and bacillary load (yellow). Displayed at left are all pathways for which any effect showed enrichment at an FDR threshold of 0.2. For each effect, pathways with an enrichment FDR < 0.2 are indicated with solid points while open points reflect FDR > = 0.2; diamond-shaped points locate two pathways uniquely enriched in the expression signal associated with cough aerosol culture positivity. Pathways are organized by their maximum enrichment score for any of the two effects. The Venn diagram at right displays the number of pathways enriched in each effect at FDR < 0.2 with stratification by concordance of directionality of effect. **d** Leading edge genes driving enrichment of Hallmark pathways uniquely associated with cough aerosol culture-positive status. Points indicate the adjusted fold change associated with cough aerosol culture positivity (y position is arbitrary jitter), are colored according to the sign of FC, and filled/labeled points indicate leading edge genes in the GSEA analysis. Gray violins indicate the distribution of FC values within each pathway, and a rug plot along the *x*-axis indicates the full distribution of FCs estimates for the cough aerosol culture-positive effect. Black curves indicate the running enrichment score deviation from zero along the *y*-axis. Source data are provided as a Source Data file.

features, cough characteristics (e.g., propulsive strength or frequency), and host inflammation. A balance in the host inflammatory response is needed to prevent the dissemination of Mtb (resulting in part from insufficient inflammation) while limiting tissue destruction and other complications from excessively robust inflammatory responses[45,46]. The stimulation of inflammatory pathways that regulate Mtb aerosolization and transmission may exert additional evolutionary pressure on the immune response to Mtb. To explore potential mechanisms of Mtb aerosolization, we discovered a whole blood transcriptomic signature associated with cough aerosol culture -positivity after adjustments for bacillary load. Given the timing of sample collection at the time of diagnosis, we cannot distinguish whether index case inflammatory pathways cause Mtb aerosolization or vice versa. If the inflammatory pathways associated with Mtb aerosolization are regulated by immunogenetically encoded mechanisms, then the Angiogenesis signaling pathway, which was enriched in cough aerosol culture-positive participants, may offer causal insights.

Of the 10 leading edge genes differentiating cough aerosol culture-positive versus -negative status, several are involved in the extracellular matrix (COL5A2, FGFR1, ITGAV) and cell development and proliferation (JAG1, PDGFA, FSTL1) pathways which offer potential insights into aerosolization mechanisms. For example, extracellular molecules secreted by these pathways could provide an alternative milieu surrounding the bacterium, which modulates the surface structure or metabolic state of Mtb and its subsequent capacity to transmit. Previous work in the *M. marinum* zebrafish model uncovered an important role for vascularization and angiogenesis in granuloma formation and control of bacterial dissemination[47,48]. Enzymatically modified trehalose dimycolate (TDM), an immunogenic cell wall lipid, induced angiogenesis through a VEGF pathway. These studies highlight potential mechanisms that could influence Mtb aerosolization via modulation of angiogenesis-dependent inflammatory pathways. Similarly, cytokine pathways (e.g., Interferon Alpha Response and Interferon Gamma Response gene sets enriched in cough aerosol culture-negative individuals) from activated immune cells could modulate Mtb's state and survival before, during, and after propulsion from the host airway. Regardless of the causal pathway, the inflammatory profiles provide a potential biomarker of infectiousness that could rapidly identify Mtb superspreaders.

Our study supports prior findings of significant individual host variation in infectiousness among patients with TB and suggests that factors beyond bacillary burden determine infectiousness[19,21,22,25,26,49]. While CASS, modeled on cough aerosol production, is the best-studied method for assessing infectiousness[50], recent studies have drawn attention to detectable Mtb from non-cough respiratory maneuvers[32–34,51]. Williams and colleagues demonstrated that when persons with confirmed and suspected TB wore face masks with collection strips (face mask sampling or FMS) while breathing normally, Mtb DNA was detected in up to 90% of persons[33]. The authors subsequently found a modest association between FMS detection of Mtb and increased TB infection in close contacts[52]. RASC is a device to capture bioaerosols that are then evaluated for viable Mtb using a fluorescent trehalose analog[32,34]. In a study of 38 GeneXpert-positive participants who performed respiratory maneuvers (tidal breathing, voluntary cough, forced vital capacity) while seated in the RASC, 88% of participants produced at least one sample positive for Mtb, and all three maneuvers were equally likely to produce viable Mtb[34]. Further investigations are needed to determine differences in findings from CASS, FMS, and RASC, especially as they relate to actual transmission events.

Our study has several limitations. First, we did not evaluate whether Mtb microbiologic factors are associated with cough aerosol culture status. A previous investigation of Mtb genetic variants and cough aerosol culture positivity did not demonstrate associations[29]. Second, 95% of participants reported cough, and our findings may not apply to persons with TB without cough. Recent studies that have detected TB in exhaled breath call into question whether cough is essential and/or a primary driver of TB transmission.[33,34,52,53]. Third, our transcriptomic data derives from whole blood which precludes cell-specific insights. Despite this limitation, pathway analysis suggests possible cellular sources of mechanisms that can be tested in future studies. Fourth, we were unable to assess the causality of immunologic pathways associated with cough aerosol culture positivity which may precede Mtb aerosolization or be a consequence of it. However, we did evaluate for possible confounding and adjusted for bacillary load to identify gene sets that are independently associated with cough aerosol culture status. Fifth, sputum collection methods are not standardized and have more technical heterogeneity in comparison to other biological samples. Despite this challenge, sputum provides a direct assessment of the primary site of TB disease, where pulmonary immune responses are compartmentalized and differ from blood. In addition, sputum cytokines have been evaluated as biomarkers in the diagnosis of TB and treatment monitoring[54–56]. However, to our knowledge, no prior studies have evaluated their association with infectiousness measured by culturable aerosols. In addition, the index participants who contributed to our evaluations of risk factors for HHC IGRA results differed from index participants without enrolled HHCs, which introduces selection bias and limits generalizability. Although our random effects models should address concerns around selection bias[57], the generalizability of the HHC IGRA status remains a limitation. Finally, we enrolled participants during the COVID-19 pandemic, and the impact of SARC-CoV-2 on CRP level, transcriptional profiles, or other study variables is not known.

There were also several strengths to our study. First, all participants underwent CASS procedures prior to the initiation of antituberculosis therapy, which is known to rapidly impact cough aerosol culture results[29,58]. Second, we adhered to a rigorous definition of cough aerosol culture status in which we excluded participants who had more than two CASS plates contaminated with overgrowth. Third, we demonstrated that cough aerosol culture-positive status was strongly associated with evidence of TB transmission in household contacts based on QFT results. The annual risk of TB infection differs by age group in high-burden settings[59], and notably, there was no difference in the ages of HHCs by CAC status. Among household

contacts less than 10 years of age, who have a lower annual risk of infection compared to adolescents and adults, QFT-positive results were more common (p-value 0.01) among contacts of CAC-positive persons (9 of 10 participants, 90%) than contacts of CAC-negative persons (14 of 33 participants, 42%).

Our findings, along with those from prior studies, suggest that host characteristics and biomarkers may identify the most infectious patients with TB. CRP, already recommended as a screening test for TB by the World Health Organization, may have a role in identifying highly infectious persons with TB[60,61]. While persons with pulmonary TB who are cough aerosol culture-negative patients may transmit Mtb, identifying the most infectious persons would allow targeted interventions to support TB control efforts, such as isolation, true direct observation of treatment, and drug-susceptibility testing to confirm that treatment is effective. TB control could also be supported through enhanced investigations to identify contacts of the most infectious persons to evaluate for active TB and provide TB preventive therapy given the higher likelihood of recent transmission and progression to disease. Further elucidation of inflammatory transcriptional signatures associated with TB infectiousness is important not only for understanding mechanisms and developing new therapies but also for the possibility of developing a diagnostic tool to identify individuals who are the most infectious.

Methods

Ethical approvals

This study complies with all relevant ethical regulations and was approved by the Kenya Medical Research Institute Scientific and Ethics Review Unit (048/3988), the Kenyatta National Hospital/University of Nairobi Institutional Review Board (KNH-ERC/A/375), and the University of Washington Institutional Review Board (STUDY00009209). All participants 18 years and older provided informed consent. For participants less than 18 years, informed consent was obtained from the legal guardian and participants ages 13 to 18 years also provided informed assent. Participants were compensated the equivalent of $4 per study visit for their time.

Study design

Study setting & participants. Between March 1, 2021, and March 30, 2023, we enrolled adults with newly diagnosed pulmonary TB in

Nairobi, Kenya. The TB incidence rate in Kenya was estimated to be 558 per 100,000 adults in 2015[62]. Participants were enrolled either through outpatient TB and respiratory clinics where they had presented for healthcare (passive case finding) as a convenience sample or were diagnosed with pulmonary TB through a Nairobi-based household TB prevalence survey (active case finding). Participants identified through active case finding underwent chest x-ray (regardless of symptoms) and cough assessment; those with an abnormal chest x-ray and/or who reported a current cough were asked to provide two sputum samples for AFB-smear, -culture, and GeneXpert testing. All participants were diagnosed with TB based on a positive GeneXpert test result, either GeneXpert MTB/RIF (Xpert MTB/RIF) or GeneXpert Ultra (Xpert Ultra). AFB culture was subsequently performed on sputum samples. A final diagnosis of pulmonary TB was based on a positive GeneXpert test unless the result was "trace positive" from the Xpert Ultra assay, in which case culture confirmation of Mtb was required. We did not include persons in this study identified through active case finding who were GeneXpert negative/culture-positive due to delays in their TB diagnosis, who were trace positive/culture-negative as we concluded that they did not have active pulmonary based on four negative sputum cultures, and those who were diagnosed on weekends or declined enrollment. Participants had not initiated anti-TB treatment at the time of enrollment and study interventions. Potential participants who were unable to consent in the study languages (Ki-Swahili, English), did not provide a home location, planned to move from the area within six months, declined study procedures, or were currently imprisoned were not eligible for the study. We also enrolled the household contacts of participants with pulmonary TB to assess for evidence of TB transmission events. Household contacts were eligible for enrollment if they resided and slept in the household for at least 60 days prior to enrollment. There were no age restrictions on the eligibility of household contacts. Participants' sex and race were recorded based on self-identification. Study data were collected and managed using REDCap electronic data capture tools hosted at the University of Washington[63].

We estimated the frequency of positive IGRAs in Nairobi residents who were not known to have active TB or had recent close contact with a person with active pulmonary TB. Participants were enrolled in two studies. As part of a study of cough analysis for TB detection[64], we recruited adult outpatients ($n = 45$) from the same TB and respiratory clinics as TBAIT participants with TB (described above) who were identified through passive case finding and presented with cough. All participants underwent sputum GeneXpert testing and chest X-rays to exclude pulmonary TB. In addition, we report results from a study to evaluate treatment responses during isoniazid preventive therapy[65]. From December 2019 to December 2020, we recruited adults with ($n = 121$) and without ($n = 122$) HIV from three HIV prevention and care centers in Nairobi. We assessed for active TB using the WHO four-symptom screen, followed by sputum GeneXpert and chest X-rays among those with one or more symptoms. To calculate the prevalence of QFT positive tests, we excluded indeterminate results.

Study procedures

Participants with TB. We collected sputa from participants at enrollment ("spot") and the following morning ("morning"). AFB-smear and culture were performed on both samples, and GeneXpert testing was performed on the spot sample. We collected separate sputum for cytokine testing in addition to whole blood for laboratory testing and mRNA analysis.

CASS consists of a six-stage Andersen Cascade Impactor (Thermo Fischer Scientific, Rockford, IL) within a larger stainless-steel chamber attached to the tubing and with a vacuum pump creating an airflow through the system[23,25]. We calibrated the vacuum pump flow rate (28.3 L/min) using a primary flow calibrator (Model 4046, TSI, Inc., Shoreview, MN) and then maintained and monitored that using a marked field rotameter (SKC, Inc., Eighty Four, PA). Each of the six stages of the Andersen cascade impactor holds a Middlebrook 7H10 or 7H11 solid agar plate, on which aerosolized particles impact based on particle size. The agar plates were loaded into the Andersen impactors in a sterile fashion; the impactors were loaded into the cylinder before each study. Prior to cough peak flow measurements and sputum collection, participants coughed into a mouthpiece on tubing connected to the chamber for 5 min. Plates were incubated at 37 °C for 8 weeks and observed weekly for growth. When growth was identified as Mtb, colony-forming units (CFU) were counted. Agar plates with non-acid fast bacilli growth were considered contaminated and discarded. The cylinder and components were autoclaved after each use. Mtb growth was first detected at a median of 4 weeks (interquartile range (IQR), 4–6) and was most frequently detected on plates four (22 with growth), five[29], and six[19], corresponding to particle sizes of 2.1–3.3 μm, 1.1–2.0 μm, and 0.65–1.1 μm, respectively[23]. (Supplementary Fig. 1) Among cough aerosol culture participants, the maximal plate CFUs, defined for each participant as the highest CFU count for any plate with Mtb growth, ranged from 1 to 76 with a median of 12 CFUs (IQR, 4–27).

For cough peak flow (CPF) measurements, we instructed participants to cough as forcefully as possible through a disposable mouthpiece attached to a Vitalograph peak flow meter (Ennis, Ireland). The procedure was performed three times, and results were recorded in L/min. The highest recorded value was used in analyses. Enrollment posteroanterior (PA) chest x-rays were obtained and interpreted by a

member of the study team (DJH, a pulmonologist) for the presence of cavitations and number of quadrants with changes attributed to TB disease; DJH was blinded to CASS results in the review of chest x-rays. Study personnel interviewed participants to collect demographic information, HIV history, and TB history. To assess cough-related effects on quality of life, we administered the Leicester Cough Questionnaire (LCQ), a validated health status measure for adults with chronic cough ranging from 3–21 with a lower score indicating greater impairment[66].

Household contacts. Consenting household contacts were interviewed to assess for symptoms consistent with TB and to collect demographic information and exposure history. Household contacts underwent chest x-ray (PA unless age < 10 years, in which case PA and lateral images were obtained) and phlebotomy for interferon-gamma release assay and HIV testing. Household contacts who were suspected of having TB based on an abnormal chest x-ray and/or the presence of symptoms (cough, fever, weight loss) had sputum collected for GeneXpert testing.

Laboratory assays. Sputum volume and quality were visually assessed by laboratory personnel. We performed concentrated sputum smears on all samples using Auramine O staining with fluorescence smear microscopy, read according to the WHO AFB scale. Sputum was processed using the NALC-NaOH-NaCitrate method, and 0.5 ml of the pellet inoculated in mycobacterial culture (MGIT) Mycobacterium Growth Indicator 4 mL tubes (Becton-Dickinson, Franklin Lakes, NJ) supplemented with BD BBL™ MGIT™ PANTA and incubated in an automated BACTEC MGIT 960 machine for growth determination. Samples that had zero(0) growth units at 42 days were confirmed negative. Broth cultures that were flagged as positive by the MGIT 960 had the time to detection recorded, and the presence of acid-fast bacilli was verified using Ziehl-Neelsen smear microscopy with isolates identified as Mtb using the MGIT TBc Identification Test (Becton-Dickinson Diagnostic Instrument Systems, Sparks, MD). Further speciation was not performed. Xpert MTB/RIF or Xpert Ultra (Cepheid, Sunnyvale, CA) were performed on raw sputum samples. GeneXpert results were recorded as a cycle threshold (Ct) value and as a semi-quantitative grade: Negative, Very Low, Low, Medium, and High, with an additional category of "Trace" for Xpert Ultra. Results from GeneXpert testing for rifampin resistance were recorded as negative, positive, or indeterminate. For the CASS culture plates, after the preparation and autoclaving of 7H10 or 7H11 solid media, appropriate volumes of OADC (Oleic acid Albumin Dextrose Catalase) and antibiotics (amphotericin B, carbenicillin, polymixin B, and trimethoprim lactate) were added to inhibit contaminant growth.

C-reactive protein (CRP) was measured from serum samples using a Cobas C 111 chemistry analyzer (Roche Diagnostics Ltd, Liechtenstein, Switzerland) according to the manufacturer's instructions. The detection range was 0.6–350 mg/L. We offered HIV testing to participants unless a participant was known to be living with HIV or if testing had been performed within the six months prior to enrollment. Whole blood for QuantiFERON-TB Plus testing (QFT-Plus, Qiagen Diagnostics; Hamburg, Germany) was drawn into lithium heparin blood collection tubes, and 1 mL amounts were transferred into Nil, Mitogen, TB1, and TB2 tubes. All QFT-Plus tubes were incubated at 37 °C within 6 h of collection. QFT-Plus processing and interpretation were performed according to the manufacturer's instructions[67]. If TB1-nil and/or TB2-nil were > 0.35 IU/ml and > 25% of nil value (with Nil < 8.0 IU/ml), then the QFT-Plus test was considered positive. If Nil > 8.0 IU/mL or Mitogen−Nil < 0.5 IU/ml with Nil < 8.0 IU/ml and negative antigen-Nil results, then QFT-Plus was defined as indeterminate. QFT-Plus results were negative if Nil < 8.0 IU/ml with either antigen-Nil values < 0.35 IU/ml IFN-γ or < 25% of Nil). Indeterminate results were repeated, with the second result reported.

Sputum and plasma cytokines. Four cytokines were examined in sputum and blood. TNF, IL6, IL1B, and CXCL8 were chosen based on their central roles in regulating inflammatory pathways and TB pathogenesis[68]. Spot sputum was digested by adding an equal volume of 10% Sputolysin (Millipore, Merck KGaA, Darmstadt, Germany). The mixture was vortexed and incubated at 37 °C for 15 min. The digested sample was centrifuged at 500 × *g* for 10 min. The supernatant was preserved by the addition of 40 µl of 25X cOmplete™ protease inhibitor cocktail solution (Roche, Merck KGaA, Darmstadt, Germany) per 1 ml of sample and stored at − 80 °C. Plasma was tested directly. Sputum supernatants and plasma were tested for CXCL8 (IL-8), Interleukin 1 beta (IL-1β), Interleukin 6 (1L-6), and TNF using sandwich ELISA according to the manufacturer's instructions. (R&D Systems Inc. Minneapolis, USA.). In pilot sputum cytokine studies, TNF was not detectable at high levels and was not examined further. In pilot plasma cytokine studies, CXCL8 and IL1B were not detectable at high levels and were not examined further.

Whole blood transcriptomics: RNA isolation, RNASeq, data processing and analysis

PAXgene tubes were thawed at room temperature, and RNA was isolated using PAXgene miRNA spin columns (Qiagen), followed by globin reduction using GlobinClear Human (ThermoFisher). To generate sequencing libraries, total RNA (0.5 ng) was added to the reaction buffer from the SMART-Seq v4 Ultra Low Input RNA Kit for Sequencing (Takara), and reverse transcription was performed, followed by PCR amplification to generate full-length amplified cDNA. Sequencing libraries were constructed using the NexteraXT DNA sample preparation kit (Illumina) to generate Illumina-compatible barcoded libraries. Libraries were pooled and quantified using a Qubit® Fluorometer (Life Technologies). Sequencing of pooled libraries was carried out on a NextSeq 2000 sequencer (Illumina) with paired-end 59-base reads, using a NextSeq P3 sequencing kit (Illumina) with a target depth of 5 million reads per sample. Base calls were processed to FASTQs on BaseSpace (Illumina), and a base call quality-trimming step was applied to remove low-confidence base calls from the ends of reads. The FASTQs were aligned to the GRCh38 human reference genome using STAR v.2.4.2a, and gene counts were generated using htseq-count. QC and metrics analysis was performed using the Picard family of tools (v1.134). Counts were assigned to gene exons using RSEM 1.3.0. Further RNA sequencing data filtering and analysis were performed in R 4.2.3[69]. We removed libraries with fewer than 1,000,000 total reads or a median coefficient of variation of coverage (median CV) greater than 0.8; this resulted in the removal of 2 sequencing libraries, both of which were clear outliers by these metrics. Libraries were reduced to protein-coding genes, and principal component analysis (PCA) did not detect any strong outliers based on RNA composition. Protein-coding counts were normalized using the trimmed mean of M-values normalization method, filtered to protein-coding genes with at least 5% of libraries containing at least 1 count per million (CPM), and finally, converted to log2 CPM using the voom-normalization procedure within the R package "limma"[70].

Data analysis

Host characteristics and household contact analyses. Our primary outcome of interest was cough aerosol culture status. Participants were considered cough aerosol culture-positive if one or more of the six CASS plates were positive for Mtb growth. Participants were considered cough aerosol culture-negative if no CASS plates were positive for Mtb growth and no more than two of the six plates were discarded as contaminated with fungal or bacterial overgrowth. Participant characteristics were compared using the chi-square test, Student's *t* test, or Wilcoxon rank-sum test as appropriate. We assessed associations between predictors and outcomes using bivariate and multivariable logistic regression. Multivariable models were developed

using forward selection stepwise regression, evaluating variables with p-values ≤ 0.20 in bivariate analyses and retaining covariates which significantly improved model fit based on likelihood-ratio tests. We assessed multicollinearity between our independent variables using variance inflation factors and condition indices. All statistical tests were two-sided with $\alpha = 0.05$.

We assessed associations between host TB characteristics and cough aerosol culture-positive status using bivariate and multivariable logistic regression models in which we evaluated predictor variables based on our conceptual model presented as a directed acyclic graph (DAG, Supplementary Fig. 2)[71]. Based on published studies of TB transmission and cough aerosol status[18,25,29], we categorized as predictors characteristics that we hypothesized as determining the cough aerosol culture-positive phenotype including younger age, functional status (body mass index (BMI), mid-upper arm circumference (MUAC), TB symptoms, cough duration, Leicester Cough Questionnaire (LCQ) score, prior history of TB, and cough peak flow)[29] and systemic inflammation (CRP, hemoglobin, white blood cell count and differential, hemoglobin A1C percent, and sputum appearance). In the DAG, we categorized HIV status, bacillary burden (GeneXpert semi-quantitative grade, GeneXpert Ct value, AFB-smear grade, time to detection (TTD) of Mtb growth in liquid culture, presence of chest X-ray cavitation, and number of radiographic quadrants with TB-related changes), and sex as potential confounders due to their shared associations with the exposure and outcome[72]. Healthcare engagement is considered a mediator, a node that is on the causal pathway from exposure to outcome, and should not be controlled in models. Pathogen-related factors, which are not included in this study, would be considered effect modifiers as they may contribute to the outcome and modify the effect of other causes of the outcome. GeneXpert Ct values were assigned using the smallest Ct value from any of the probes targeting the rpoB gene; participants with Xpert Ultra trace positive results (for whom rpoB probe Ct values were 0) were assigned a Ct value of 35, near the highest detectable Ct value for Xpert Ultra.

To evaluate associations between cough aerosol culture status and evidence of TB transmission, we investigated QFT-Plus responses in household contacts by cough aerosol culture status of the index participant. QFT-Plus responses were dichotomized as positive (≥ 0.35 IU/mL) or negative (< 0.35 IU/mL) after excluding indeterminate responses. We also evaluated the absolute interferon-gamma response based on the larger value of either TB antigen 1 − nil or TB antigen 2 − nil. To evaluate associations with QuantiFERON positive results in household contacts, we used bivariate and multivariable logistic regression models with random intercepts (*melogit* command in Stata) to account for clustering by index participant and compared multivariable models with the likelihood-ratio test. and R: A Language and Environment for Statistical Computing.

Risk score modeling

Analyses were performed to develop a risk score for clinical decision-making to identify highly infectious (cough aerosol culture-positive) persons with pulmonary TB. We evaluated covariates from our multivariable logistic regression model that could be applied at the time of a patient's diagnosis with pulmonary TB. Included variables were categorized, and the optimal cut-points were determined using the Youden index (J) method, the point maximizing the Youden function, which is the difference between true positive rate and false positive rate over all possible cut-point values[73]. These values were then used in a multivariable logistic regression model. The risk score was generated by dividing the beta coefficients from the logistic regression model by the smallest beta in the model, multiplying by 5, and rounding to the nearest integer. Internal validation was conducted using the bootstrap resampling method with 5000 replications. We evaluated Somers' D_{xy} index, a rank correlation between predicted probabilities and observed responses, to evaluate model performance[74]. Somers' D_{xy} index takes values between −1 and 1, with the latter demonstrating agreement between predicted and observed responses.

Cytokine analyses

Association of cytokine concentrations (log10 pg/ml) with cough aerosol culture status and cavitary disease was done by Wilcoxon rank sum test with continuity correction. Bivariate logistic regression models were used to compare cytokine concentration with cough aerosol culture status, GeneXpert cycle thresholds, and number of chest x-ray quadrants with changes attributed to TB disease. In multivariate analyses, we evaluated associations between the outcome of cough aerosol culture status and cytokine concentrations adjusting for age, sex, cavitations on chest x-ray, and GeneXpert cycle threshold.

RNASeq Analysis: estimation of differential expression

Fold changes in gene expression were estimated using linear models implemented in the R package "kimma".[75]. In addition to the primary fixed effect of cough aerosol culture, model selection considered for inclusion of clinical covariates that were significantly associated with cough aerosol culture status in simple bivariate modeling, and which were complete for our RNASeq subset. We compared single covariate additions of bacillary load (GeneXpert Ct), age, and cavitary disease, assessing trends in model fit using the Akaike Information Criterion (AIC) across the genome (Supplementary Fig. 4a). We then performed stepwise model addition, comparing the impact on AIC at each step, and only retained covariates which significantly improved model fit (delta AIC < − 2) for at least 10% of genes (Supplementary Fig. 4B). This resulted in an expression model accounting for bacillary load in addition to the cough aerosol culture phenotype (Supplementary Fig. 4c, d).

RNASeq Analysis: gene set enrichment

We used a competitive gene set test to evaluate the enrichment of gene expression differences between cough aerosol culture status (positive vs. negative). We used gene set enrichment analysis (GSEA) using the pre-ranked approach employed in the fast GSEA (FGSEA) method and implemented in the R package "fgsea"[76], and tested for enrichment of the Hallmark annotated biological pathways accessed via the mSigDB database[77]. Log2 Fold change estimates (log2 FC) for cough aerosol culture status, adjusted for the effect of bacillary load, were used to pre-rank genes for GSEA. Estimates for the effects of bacillary load were separately tested for enrichment in GSEA to identify enriched pathways unique to cough aerosol culture status. For each clinical predictor (cough aerosol culture status, bacillary load), we used kimma estimated log2FC to pre-rank genes and used gene-set permutation with 1000 permutations to randomize gene ordering[75].

Reporting summary

Further information on research design is available in the Nature Portfolio Reporting Summary linked to this article.

Data availability

The RNA-seq data generated in this study have been deposited in the dbGaP database (Home-dbGaP-NCBI(nih.gov)) under accession code phs003727.v1.p1. Source data are provided in this paper.

References

1. World Health Organization. Global Tuberculosis Report 2017. Geneva, Switzerland: http://apps.who.int/iris/bitstream/10665/259366/1/9789241565516-eng.pdf?ua=1 (2017).
2. Yuen, C. M. et al. Turning off the tap: stopping tuberculosis transmission through active case-finding and prompt effective treatment. *Lancet* **386**, 2334–2343 (2015).

3. Auld, S. C. et al. Research roadmap for tuberculosis transmission science: Where do we go from here and how will we know when we're there? *J. Infect. Dis.* **216**, S662–S668 (2017).
4. Woolhouse, M. E. et al. Heterogeneities in the transmission of infectious agents: implications for the design of control programs. *Proc. Natl. Acad. Sci. USA* **94**, 338–342 (1997).
5. Smith, D. L., Dushoff, J., Snow, R. W. & Hay, S. I. The entomological inoculation rate and Plasmodium falciparum infection in African children. *Nature* **438**, 492–495 (2005).
6. Teicher, A. Super-spreaders: a historical review. *Lancet Infect. Dis.* **23**, e409–e417 (2023).
7. Ypma, R. J., Altes, H. K., van Soolingen, D., Wallinga, J. & van Ballegooijen, W. M. A sign of superspreading in tuberculosis: highly skewed distribution of genotypic cluster sizes. *Epidemiology* **24**, 395–400 (2013).
8. McCreesh, N. & White, R. G. An explanation for the low proportion of tuberculosis that results from transmission between household and known social contacts. *Sci. Rep.* **8**, 5382 (2018).
9. Snider, D. E. Jr., Kelly, G. D., Cauthen, G. M., Thompson, N. J. & Kilburn, J. O. Infection and disease among contacts of tuberculosis cases with drug-resistant and drug-susceptible bacilli. *Am. Rev. Respir. Dis.* **132**, 125–132 (1985).
10. Kline, S. E., Hedemark, L. L. & Davies, S. F. Outbreak of tuberculosis among regular patrons of a neighborhood bar. *N. Engl. J. Med.* **333**, 222–227 (1995).
11. Lee, R. S. et al. Reemergence and amplification of tuberculosis in the Canadian arctic. *J. Infect. Dis.* **211**, 1905–1914 (2015).
12. Lloyd-Smith, J. O., Schreiber, S. J., Kopp, P. E. & Getz, W. M. Superspreading and the effect of individual variation on disease emergence. *Nature* **438**, 355–359 (2005).
13. Melsew, Y. A. et al. Risk factors for infectiousness of patients with tuberculosis: a systematic review and meta-analysis. *Epidemiol. Infect.* **146**, 345–353 (2018).
14. Turner, R. D. & Bothamley, G. H. Cough and the transmission of tuberculosis. *J. Infect. Dis.* **211**, 1367–1372 (2015).
15. Donald, P. R. et al. Droplets, dust and guinea pigs: an historical review of tuberculosis transmission research, 1878-1940. *Int J. Tuberc. Lung Dis.* **22**, 972–982 (2018).
16. van Geuns, H. A., Meijer, J. & Styblo, K. Results of contact examination in Rotterdam, 1967-1969. *Bull. Int Union Tuberc.* **50**, 107–121 (1975).
17. Zurcher, K. et al. Estimating tuberculosis transmission risks in a primary care clinic in South Africa: modeling of environmental and clinical data. *J. Infect. Dis.* **225**, 1642–1652 (2022).
18. Borgdorff, M. W., Nagelkerke, N. J., de Haas, P. E. & van Soolingen, D. Transmission of Mycobacterium tuberculosis depending on the age and sex of source cases. *Am. J. Epidemiol.* **154**, 934–943 (2001).
19. Riley, R. L., Wells, W. F., Mills, C. C., Nyka, W. & McLean, R. L. Air hygiene in tuberculosis: quantitative studies of infectivity and control in a pilot ward. *Am. Rev. Tuberc.* **75**, 420–431 (1957).
20. Sultan, L. et al. Tuberculosis disseminators. A study of the variability of aerial infectivity of tuberculous patients. *Am. Rev. Respir. Dis.* **82**, 358–369 (1960).
21. Riley, R. L. et al. Aerial dissemination of pulmonary tuberculosis. A two-year study of contagion in a tuberculosis ward. 1959. *Am. J. Epidemiol.* **142**, 3–14 (1995).
22. Escombe, A. R. et al. The infectiousness of tuberculosis patients coinfected with HIV. *PLoS Med.* **5**, e188 (2008).
23. Fennelly, K. P. et al. Cough-generated aerosols of Mycobacterium tuberculosis: a new method to study infectiousness. *Am. J. Respir. Crit. Care Med.* **169**, 604–609 (2004).
24. Acuna-Villaorduna, C. et al. Cough-aerosol cultures of Mycobacterium tuberculosis in the prediction of outcomes after exposure. A household contact study in Brazil. *PLoS ONE* **13**, e0206384 (2018).
25. Fennelly, K. P. et al. Variability of infectious aerosols produced during coughing by patients with pulmonary tuberculosis. *Am. J. Respir. Crit. Care Med.* **186**, 450–457 (2012).
26. Jones-Lopez, E. C. et al. Cough aerosols of mycobacterium tuberculosis in the prediction of incident tuberculosis disease in household contacts. *Clin. Infect. Dis.* **63**, 10–20 (2016).
27. Jones-Lopez, E. C. et al. Cough aerosols of Mycobacterium tuberculosis predict new infection: a household contact study. *Am. J. Respir. Crit. Care Med.* **187**, 1007–1015 (2013).
28. Jones-Lopez, E. C. et al. Cough aerosol cultures of mycobacterium tuberculosis: Insights on TST / IGRA discordance and transmission dynamics. *PLoS ONE* **10**, e0138358 (2015).
29. Theron, G. et al. Bacterial and host determinants of cough aerosol culture positivity in patients with drug-resistant versus drug-susceptible tuberculosis. *Nat. Med.* **26**, 1435–1443 (2020).
30. Wood, R. et al. Real-time investigation of tuberculosis transmission: Developing the respiratory aerosol sampling chamber (RASC). *PLoS ONE* **11**, e0146658 (2016).
31. Patterson, B. et al. Bioaerosol sampling of patients with suspected pulmonary tuberculosis: a study protocol. *BMC Infect. Dis.* **20**, 587 (2020).
32. Dinkele, R. et al. Capture and visualization of live Mycobacterium tuberculosis bacilli from tuberculosis patient bioaerosols. *PLoS Pathog.* **17**, e1009262 (2021).
33. Williams, C. M. et al. Exhaled Mycobacterium tuberculosis output and detection of subclinical disease by face-mask sampling: prospective observational studies. *Lancet Infect. Dis.* **20**, 607–617 (2020).
34. Dinkele, R. et al. Aerosolization of Mycobacterium tuberculosis by Tidal Breathing. *Am. J. Respir. Crit. Care Med.* **206**, 206–216 (2022).
35. Behr, M. A., Edelstein, P. H. & Ramakrishnan, L. Revisiting the timetable of tuberculosis. *BMJ* **362**, k2738 (2018).
36. Dowdy, D. W., Azman, A. S., Kendall, E. A. & Mathema, B. Transforming the fight against tuberculosis: targeting catalysts of transmission. *Clin. Infect. Dis.* **59**, 1123–1129 (2014).
37. Gentile I., et al. The role of CRP POC testing in the fight against antibiotic overuse in European primary care: Recommendations from a European expert panel. *Diagnostics* **13**, https://doi.org/10.3390/diagnostics13020320 (2023).
38. Dhana, A. et al. Tuberculosis screening among ambulatory people living with HIV: a systematic review and individual participant data meta-analysis. *Lancet Infect. Dis.* **22**, 507–518 (2022).
39. de Hollander, E. L., Bemelmans, W. J. & de Groot, L. C. Associations between changes in anthropometric measures and mortality in old age: a role for mid-upper arm circumference? *J. Am. Med. Dir. Assoc.* **14**, 187–193 (2013).
40. Schaap, L. A., Quirke, T., Wijnhoven, H. A. H. & Visser, M. Changes in body mass index and mid-upper arm circumference in relation to all-cause mortality in older adults. *Clin. Nutr.* **37**, 2252–2259 (2018).
41. Wijnhoven, H. A. et al. Low mid-upper arm circumference, calf circumference, and body mass index and mortality in older persons. *J. Gerontol. A Biol. Sci. Med. Sci.* **65**, 1107–1114 (2010).
42. Getnet, F. et al. Delay in diagnosis of pulmonary tuberculosis increases the risk of pulmonary cavitation in pastoralist setting of Ethiopia. *BMC Pulm. Med.* **19**, 201 (2019).
43. Baazim, H., Antonio-Herrera, L. & Bergthaler, A. The interplay of immunology and cachexia in infection and cancer. *Nat. Rev. Immunol.* **22**, 309–321 (2022).
44. Redford, S. E., Varanasi, S. K., Sanchez, K. K., Thorup, N. R. & Ayres, J. S. CD4+ T cells regulate sickness-induced anorexia and fat wasting during a chronic parasitic infection. *Cell Rep.* **42**, 112814 (2023).
45. Dorhoi, A. & Kaufmann, S. H. Perspectives on host adaptation in response to Mycobacterium tuberculosis: modulation of inflammation. *Semin Immunol.* **26**, 533–542 (2014).

46. Matty, M. A., Roca, F. J., Cronan, M. R. & Tobin, D. M. Adventures within the speckled band: heterogeneity, angiogenesis, and balanced inflammation in the tuberculous granuloma. *Immunol. Rev.* **264**, 276–287 (2015).
47. Oehlers, S. H. et al. Interception of host angiogenic signalling limits mycobacterial growth. *Nature* **517**, 612–615 (2015).
48. Walton, E. M. et al. Cyclopropane modification of trehalose dimycolate drives granuloma angiogenesis and mycobacterial growth through VEGF signaling. *Cell Host Microbe* **24**, 514–25 e6 (2018).
49. Riley, R. L. et al. Infectiousness of air from a tuberculosis ward. Ultraviolet irradiation of infected air: comparative infectiousness of different patients. *Am. Rev. Respir. Dis.* **85**, 511–525 (1962).
50. Coussens A. K., et al. Classification of early tuberculosis states to guide research for improved care and prevention: an international Delphi consensus exercise. *Lancet Respir Med.* **6**, 484–498 (2024).
51. Williams, C. M. et al. Face mask sampling for the detection of Mycobacterium tuberculosis in expelled aerosols. *PLoS ONE* **9**, e104921 (2014).
52. Williams, C. M. et al. Exhaled Mycobacterium tuberculosis predicts incident infection in household contacts. *Clin. Infect. Dis.* **76**, e957–e964 (2023).
53. Patterson, B. et al. Aerosolization of viable Mycobacterium tuberculosis bacilli by tuberculosis clinic attendees independent of sputum-Xpert Ultra status. *Proc. Natl. Acad. Sci. USA* **121**, e2314813121 (2024).
54. Ribeiro-Rodrigues, R. et al. Sputum cytokine levels in patients with pulmonary tuberculosis as early markers of mycobacterial clearance. *Clin. Diagn. Lab Immunol.* **9**, 818–823 (2002).
55. Ota, M. O. et al. Rapid diagnosis of tuberculosis using ex vivo host biomarkers in sputum. *Eur. Respir. J.* **44**, 254–257 (2014).
56. Heslop, R. et al. Changes in host cytokine patterns of TB patients with different bacterial loads detected using 16S rRNA analysis. *PLoS ONE* **11**, e0168272 (2016).
57. Bell, A., Fairbrother, M. & Jones, K. Fixed and random effects models: making an informed choice. *Qual. Quant.* **53**, 1051–1074 (2019).
58. Acuna-Villaorduna, C. et al. Host determinants of infectiousness in smear-positive patients With pulmonary tuberculosis. *Open Forum Infect. Dis.* **6**, ofz184 (2019).
59. Dowdy, D. W. & Behr, M. A. Are we underestimating the annual risk of infection with Mycobacterium tuberculosis in high-burden settings? *Lancet Infect. Dis.* **22**, e271–e278 (2022).
60. Yew, W. W. & Leung, C. C. Are some people not safer after successful treatment of tuberculosis? *Am. J. Respir. Crit. Care Med.* **171**, 1324–1325 (2005).
61. Martinez-Gonzalez N. A., et al. Point-of-care C-reactive protein testing to reduce antibiotic prescribing for respiratory tract infections in primary care: Systematic review and meta-analysis of randomised controlled trials. *Antibiotics* **9**, https://doi.org/10.3390/antibiotics9090610 (2020).
62. Enos, M. et al. Kenya tuberculosis prevalence survey 2016: Challenges and opportunities of ending TB in Kenya. *PLoS ONE* **13**, e0209098 (2018).
63. Harris, P. A. et al. Research electronic data capture (REDCap)-a metadata-driven methodology and workflow process for providing translational research informatics support. *J. Biomed. Inf.* **42**, 377–381 (2009).
64. Sharma, M. et al. TBscreen: A passive cough classifier for tuberculosis screening with a controlled dataset. *Sci. Adv.* **10**, eadi0282 (2024).
65. Njagi L. N., Nduba V., Mureithi M. W., Mecha J. O. Prevalence and predictors of tuberculosis infection among people living with HIV in a high tuberculosis burden context. *BMJ Open Respir Res.* https://doi.org/10.1136/bmjresp-2022-001581 (2023).
66. Birring, S. S. et al. Development of a symptom specific health status measure for patients with chronic cough: Leicester Cough Questionnaire (LCQ). *Thorax* **58**, 339–343 (2003).
67. Hopewell, P. C., Pai, M., Maher, D., Uplekar, M. & Raviglione, M. C. International standards for tuberculosis care. *Lancet Infect. Dis.* **6**, 710–725 (2006).
68. Mayer-Barber K. D., Barber D. L. Innate and Adaptive Cellular Immune Responses to Mycobacterium tuberculosis Infection. *Cold Spring Harb Perspect Med.* **5**, a018424 (2015).
69. R. Core Team. R: A language and environment for statistical computing. 4.2.3 ed: (Foundation for Statistical Computing, Vienna, Austria; 2023).
70. Law, C. W., Chen, Y., Shi, W. & Smyth, G. K. voom: Precision weights unlock linear model analysis tools for RNA-seq read counts. *Genome Biol.* **15**, R29 (2014).
71. Greenland, S., Pearl, J. & Robins, J. M. Causal diagrams for epidemiologic research. *Epidemiology* **10**, 37–48 (1999).
72. Digitale, J. C., Martin, J. N. & Glymour, M. M. Tutorial on directed acyclic graphs. *J. Clin. Epidemiol.* **142**, 264–267 (2022).
73. Youden, W. J. Index for rating diagnostic tests. *Cancer* **3**, 32–35 (1950).
74. Somers, R. H. A new asymmetric measure of association for ordinal variables. *Am. Sociol. Rev.* **27**, 799–811 (1962).
75. Dill-McFarland K. A., et al. Kimma: flexible linear mixed effects modeling with kinship covariance for RNA-seq data. Bioinformatics. **39**, btad279 (2023).
76. Korotkevich G., et al. Fast gene set enrichment analysis. Preprint at bioRxiv https://doi.org/10.1101/060012 (2021).
77. Subramanian, A. et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. USA* **102**, 15545–15550 (2005).

Acknowledgements

This research was funded by the National Institutes of Health/NIAID (NIH grant 5R01AI150815 - DJH/TRH/VN and UH2AI152621 - DJH), NIH D43 TW011817-01 (VN/LNN/WB/DJH/TRH), and the Firland Foundation (TRH). The University of Washington/Fred Hutch Center for AIDS Research (AI027757) provided support to the J.B. REDCap database for data collection was supported by the National Center for Advancing Translational Sciences of the NIH (UL1 TR002319). K.F. was funded entirely by the NHLBI Division of Intramural Research. We would like to thank the participation of the individual study participants and their families. We thank Dr. Lucy Kijaro, Dr. Joy Githua, Dr. Jacqueline Mirera, Robi Chacha, Lenis Njagi, Geoffrey Onchiri, Ruth Munyasya, Caroline Epiche, Isaac Kibet, Stella Nthambi, Kevin Munge, Japherson Mecha, Patrick Isinidu, Joash Omolo, Hastings Koech, Inviolata Sakwa and all other KEMRI CRDR Nairobi staff for their support in data collection. We thank Madison Jones for laboratory assistance.

Author contributions

V.N., K.F., T.R.H., and D.J.H. designed the study. L.N.N., V.N., G.L., T.R.H., and D.J.H. led participant enrollment and study interventions. K.F. supervised CASS studies. W.M., Z.M., and G.P. led laboratory investigations under the supervision of T.R.H. and V.N. Statistical analyses were performed by D.J.H., R.M.S, and J.B. D.J.H. and T.R.H. drafted the manuscript, and all authors edited and reviewed the final manuscript. All collaborators of this study have fulfilled the criteria for authorship required by Nature Portfolio journals and have been included as authors, as their participation was essential for the design and implementation of the study. Roles and responsibilities were agreed among collaborators ahead of the research. This work includes findings that are locally relevant; this was determined in collaboration with local partners. We obtained local ethics reviews of all research in this paper. This research was not severely restricted or prohibited in the setting of the researchers and did not result in stigmatization, discrimination or personal risk to

participants. When available, we included local and regional research relevant to our study in citations.

Competing interests
The authors declare no competing interests.

Additional information
Supplementary information The online version contains supplementary material available at https://doi.org/10.1038/s41467-024-52122-x.

Correspondence and requests for materials should be addressed to David J. Horne.

Peer review information *Nature Communications* thanks Mariana Araújo-Pereira, Delia Goletti, and the other anonymous reviewer(s) for their contribution to the peer review of this work. A peer review file is available.