Comparative assessment of line probe assays and targeted next-generation sequencing in drug-resistant tuberculosis diagnosis

Giovanna Carpi,[a] Marva Seifert,[a,b] Andres De la Rossa,[a] Swapna Uplekar,[a] Camilla Rodrigues,[c] Nestani Tukvadze,[d,e] Shaheed V. Omar,[f] Anita Suresh,[a] Timothy C. Rodwell,[a,b] and Rebecca E. Colman[a,b,*]

[a]FIND, Geneva, Switzerland
[b]Division of Pulmonary, Critical Care, Sleep Medicine, and Physiology, University of San Diego, San Diego, CA, USA
[c]Hinduja Hospital and Medical Research Centre, Tuberculosis, Mumbai, India
[d]National Center for Tuberculosis and Lung Diseases, Tuberculosis, Tbilisi, Georgia
[e]Swiss Tropical and Public Health Institute, Allschwil, Switzerland
[f]Centre for Tuberculosis, National & WHO Supranational TB Reference Laboratory, National Institute for Communicable Diseases, A Division of the National Health Laboratory Service, Johannesburg, South Africa

eBioMedicine
2025;119: 105875
Published Online xxx
https://doi.org/10.1016/j.ebiom.2025.105875

Summary

Background Rapid and accurate detection of drug-resistant tuberculosis (DR-TB) is crucial for ensuring effective treatment, halting transmission and preventing the amplification of resistance. Comparative evaluations of molecular diagnostic assays in high-burden settings are essential for informing clinical decision-making for DR-TB treatment.

Methods The Seq&Treat clinical study previously evaluated the performance of two targeted next-generation sequencing (tNGS) workflows, GenoScreen Deeplex Myc-TB and Oxford Nanopore Technologies Tuberculosis Drug Resistance Test, on direct sediment samples from persons at risk for DR-TB. Hain Line Probe Assay (LPAs—MTBDR*plus* and MTBDR*sl*) were run as a comparator test using an aliquot of the same sediment samples. Diagnostic performance of the LPAs and previously established tNGS performance were compared, including sensitivity and specificity, for rifampicin, isoniazid, fluoroquinolones (moxifloxacin, levofloxacin), and amikacin, using a composite reference standard of phenotypic drug susceptibility testing and whole-genome sequencing.

Findings Among 720 clinical samples tested, MTBDR*plus* LPA sensitivity for rifampicin and isoniazid was 92.3% (95% CI 88.9–94.8) and 91.9% (88.4–94.4), each significantly lower than ≥95% achieved by both tNGS workflows ($p < 0.01$). For fluoroquinolones (moxifloxacin and levofloxacin), the MTBDR*sl* LPA and ONT had similar sensitivities (94.3% and 92.7%, and 94.8% and 93.9%, respectively), while GenoScreen outperformed both (97.3% and 96.6%). GenoScreen also demonstrated the highest sensitivity for amikacin resistance (94.6%) compared to LPAs (88.7%) and ONT (88.3%). Complete assay failure rates were low for LPAs (4.9%) and ONT (5.0%) and moderately higher for GenoScreen (8.6%), with differences in single-target failures across all assays.

Interpretation LPAs demonstrated lower sensitivity and more limited drug resistance detection compared to tNGS workflows, underscoring the advantages of tNGS for improving DR-TB diagnostic algorithms. These findings provide critical evidence to guide updates in DR-TB diagnostic programs.

Funding Support for the Seq&Treat project was provided through funding from Unitaid (2019-32-FIND MDR).

Keywords: Targeted NGS; tNGS; Sequencing; Drug resistance; Tuberculosis; Line probe assay

Introduction

Drug-resistant tuberculosis (DR-TB) poses a significant challenge to global tuberculosis control efforts. The COVID-19 pandemic has hindered access to TB diagnostic and treatment programs, reversing years of progress.[1] In 2023, an estimated 10.8 million people fell

*Corresponding author. Division of Pulmonary, Critical Care, Sleep, Medicine, and Physiology, University of San Diego, San Diego, CA 92093, USA.
E-mail addresses: rcolman@health.ucsd.edu, rcolman@ucsd.edu (R.E. Colman).

Research in context

Evidence before this study

A search conducted on PubMed on November 19, 2024, for relevant published research using the terms (line probe assay OR LPA) AND (targeted next generation sequencing OR NGS) AND tuberculosis AND (drug resistance OR drug-resistant) AND (clinical utility OR sputum OR comprehensive genetic information) yielded five studies between 2016 and the present, with only one clinical study comparing the diagnostic accuracy of LPAs and one tNGS solution. This highlights the pressing need for additional comparative performance evaluations of these diagnostic assays. While rapid molecular assays have improved case detection rates for tuberculosis drug resistance, limitations remain for detecting resistance to certain first- and second-line drugs. Growth-based phenotypic drug susceptibility testing remains the gold standard for comprehensive resistance profiling but is hampered by long turnaround times. LPAs provide a faster, cost-effective alternative, yet their sensitivity for key drugs and lack of coverage for new and repurposed treatments limit their utility. tNGS is a promising alternative that also enables direct resistance testing of clinical specimens but for an expanded drug list. However, clinical validation and comparative evaluation of commercially available LPAs and tNGS workflows, such as GenoScreen Deeplex Myc-TB and Oxford Nanopore Technologies Tuberculosis Drug Resistance Test, remain limited, underscoring the need for further research.

Added value of this study

This study reports a large-scale, multicentre clinical evaluation directly comparing Hain Genotype LPAs with two targeted NGS (tNGS) workflows—GenoScreen Deeplex Myc-TB (GS) and Oxford Nanopore Technologies Tuberculosis Drug Resistance Test (ONT)—for diagnosing drug-resistant tuberculosis (DR-TB) across diverse settings. The study provides a comprehensive assessment of diagnostic accuracy, failure rates, and comparative performance for all drug targets included in the LPAs. While LPAs demonstrate reliable drug resistance detection with low error rates, tNGS offers higher sensitivity and broader mutation coverage, particularly for complex resistance profiles, underscoring its potential role in improving diagnostic algorithms.

Implications of all the available evidence

These findings provide critical evidence for optimising DR-TB diagnostic strategies. LPAs remain valuable tools but have limitations in mutation coverage and sensitivity for key drugs. tNGS, with its broader resistance detection capabilities, is well-suited for settings requiring rapid and comprehensive profiling. This study supports efforts to refine national TB diagnostic programs, strengthen capacity for tNGS implementation, and enable more individualised treatment decisions to improve patient outcomes.

ill with TB worldwide, a 6.5% rise from 2020, and an estimated 400,000 people developed multidrug-resistant (MDR-TB) or rifampicin-resistant TB (RR-TB).[2] Yet only an estimated 44% of individuals diagnosed with MDR/RR-TB were enrolled on treatment. Additionally, although the global burden of rifampicin-susceptible but isoniazid-resistant strains is more than double that of RR-TB, these cases remain largely undetected and untreated due to limitations in current diagnostic methods.[2,3] Thus, the rapid and accurate detection of DR-TB is essential for guiding effective treatment regimens and for preventing the further spread of resistance strains.

Culture-based phenotypic drug susceptibility testing (pDST) remains the gold standard for comprehensive drug resistance profiling in clinical settings. However, pDST has significant drawbacks, including lengthy turnaround times due to the slow growth of *Mycobacterium tuberculosis* (*Mtb*), stringent quality control requirements, and the need for extensive infrastructure, which limits its feasibility in many resource-constrained settings. Molecular diagnostics offer a rapid alternative for profiling DR-TB and have enhanced global DR-TB detection, particularly for first-line drugs.[1] Line probe assays (LPAs) are molecular WHO-recommended rapid diagnostics (mWRDs) that use DNA test strips to determine the drug resistance profile of *Mtb* strains by analysing band patterns formed through the hybridisation of immobilised probes with *Mtb* amplicons. LPA probes target the most common mutations in specific *Mtb* gene regions associated with resistance to first- and second-line drugs, as well as specific wild-type (WT) sequences.[4,5] LPAs provide a rapid and cost-effective alternative to pDST. However, the sensitivity of LPAs, particularly for detecting resistance to isoniazid and second-line injectable drugs, could be improved.[6,7] Furthermore, due to the inherent design complexities of probe-based assays, LPAs only target a small number of known resistance-conferring mutations and are restricted to a specific set of drugs. They are also prone to systematic errors due to synonymous and non-synonymous mutations not associated with resistance which can affect probe binding.[5,8–10] Notably, LPAs do not include targets to detect resistance to new and repurposed drugs such as bedaquiline (BDQ) pretomanid (Pa), and linezolid (LZD), which are integral components of the 6-month BPaLM and BPaL regimens endorsed by WHO for the treatment of MDR/RR-TB and pre-extensively drug-resistant TB (pre-XDR-TB; MDR/RR-TB also resistant to any fluoroquinolones).[11]

In contrast, high-throughput amplicon sequencing-based methods, such as targeted next-generation sequencing (tNGS) workflows, offer a comprehensive approach for detecting drug resistance to multiple tuberculosis drugs simultaneously, directly from clinical samples. tNGS workflows can swiftly detect resistance to both first- and second-line tuberculosis drugs, as well as to new and repurposed drugs, surpassing the capabilities of existing mWRDs.[12–18] Various tNGS assays for DR-TB profiling, differing in number of loci, target sizes, and NGS platforms utilised, have been previously described.[14,19–21] End-to-end tNGS workflows, GenoScreen Deeplex Myc-TB,[22,23] and Oxford Nanopore Technologies Tuberculosis Drug Resistance Test, received WHO endorsement in March 2024 as new tools for guiding clinical decisions in DR-TB treatment, with conditional recommendations on their use.[24]

While the clinical diagnostic accuracy of these end-to-end tNGS workflows for comprehensive profiling of DR-TB from clinical samples in diverse epidemiological settings have been previously reported,[18] a comparative assessment of WHO-endorsed LPAs and tNGS in reference settings is still lacking. This study aimed to fill that gap. Utilising the previously published tNGS data from a cross-sectional, multicentre clinical diagnostic study,[18] we compared the performance of the GenoType MTBDR*plus* and MTBDR*sl* LPAs and the two commercial end-to-end tNGS workflows: GenoScreen Deeplex Myc-TB (GS) and Oxford Nanopore Technologies (ONT) TB Drug Resistance Test. We evaluated and compared the diagnostic accuracy of these assays in detecting resistance to rifampicin (RIF), isoniazid (INH), and fluoroquinolones (FQ; moxifloxacin (MFX), levofloxacin (LFX)), and amikacin (AMK), using a composite reference standard of culture-based phenotypic drug susceptibility testing (pDST) and whole-genome sequencing (WGS). Additionally, we analysed sample and target failure rates across these assays.

Methods

Ethics statement

The study was approved by the WHO Research Ethics Review Committee (ERC) and by participating institutions for each study site, including the Institutional Ethics Committee of Hinduja Hospital and the Health Ministry Screening Committee of the Indian Council of Medical Research (HMSC), the University of Witwatersrand Institutional Ethics Review Committee and the Nation Health Research Database of South Africa, and the Ethics Committee of the National Center for Tuberculosis and Lungs Diseases of Georgia. All study participants provided written informed consent. The study, Seq&Treat, is registered at ClinicalTrials.gov (NCT04239326). The ethical approval by WHO Research Ethics Review Committee is ERC.0003342.

Study design and sample processing

As described previously, clinical specimens were collected as part of the prospective, cross-sectional, multicentre clinical diagnostic accuracy Seq&Treat tNGS study. The study is registered at clinicaltrials.gov (NCT04239326). Detailed descriptions of the study protocol and major tNGS findings have been published elsewhere.[18] In brief, the cross-sectional clinical study was conducted at the TB reference laboratories in Georgia, India, and South Africa between 16 April 2021 and 30 June 2022. Eligible participants at risk for DR-TB with sputum confirmed positive for *Mtb* by GeneXpert MTB/RIF (Cepheid, Sunnyvale, CA, USA), were invited to participate in the Seq&Treat tNGS study at the respective clinic sites. Sputum samples (at least 6 mL) were collected, homogenised, decontaminated, and re-suspended as sediments in 4 mL final volume for all downstream testing. GeneXpert MTB/RIF, Acid-fast bacilli (AFB) smear, Mycobacteria Growth Indicator Tube (MGIT), GenoType MTBDR*plus* and MTBDR*sl* LPAs were all performed on the same sediment for standard of care and comparator testing (Supplementary Fig. S1). GeneXpert MTB/RIF was used as a proxy of bacterial load of the study sediments on which all reference standard and comparator tests were run. The GeneXpert instrument reports semi-quantitative categories as a measure of bacterial load (Cepheid package-insert Doc 301-0191 Rev E). These same semi-quantitative categories were used for all downstream analyses of assay performance.

The index testing sediment was frozen as individual aliquots with volumes depending on requirements corresponding to specific end-to-end tNGS workflows.[18] Each individual aliquot used for tNGS workflows underwent a single freeze-thaw cycle. For the LPA vs. tNGS analysis, reference MGIT DST was performed for all culture-positive samples to determine phenotypic resistance to first-line and second-line anti-tuberculous drugs at WHO endorsed critical concentrations[18,25]: rifampicin (0.5 μg/ml), isoniazid (0.1 μg/ml), moxifloxacin (0.25 μg/ml), levofloxacin (1.0 μg/ml), amikacin (1.0 μg/ml), capreomycin (2.5 μg/ml), and kanamycin (2.5 μg/ml). DNA extracted from positive culture samples were sent for Illumina WGS, and all raw sequencing data were run through the same WHO WGS pipeline[18] and interpreted as resistant or susceptible based on the 2021 WHO catalogue of *Mtb* mutations.[26,27] All diagnostic tests were conducted at the respective reference laboratories on-site, and all diagnostic test results (GeneXpert MTB/RIF, smear, MGIT, phenotypic DST, MTBDR*plus*, MTBDR*sl*, and tNGS workflows) were recorded by the laboratory technicians for each site. The LPA data has not been previously analysed nor published.

Reference standards

Phenotypic DST results were considered resistant based on growth at the WHO endorsed critical concentration

for each drug[25] (see Methods above). WGS results were considered resistant based on detection of mutations "associated with resistance" or "associated with resistance interim" as reported in the 2021 WHO Catalogue of *Mtb* mutations.[26,27] Results of pDST and reference WGS were combined to create a composite reference result of "resistant" or "susceptible". The use of a composite reference standard enables a more robust benchmark,[25] allowing for a comprehensive evaluation of diagnostic performance. Clinical *Mtb* isolates were labelled "resistant" for a specific drug if either the pDST or WGS results were resistant and "susceptible" if both pDST and WGS showed the isolate to be susceptible to the specific drug.

Line probe assays

Each LPA assay was performed on 500 µl fresh decontaminated sediment from each clinical sample using the GenoType MTBDR*plus* version 2.0 and GenoType MTBDR*sl* version 2.0 (Hain Lifescience, Nehren, Germany). LPAs were performed and interpreted according to the Hain manufacturer instructions,[5,28,29] and standard procedures. Briefly, DNA was extracted from up to 10 sediment samples and two controls simultaneously using the GenoLyse kit (Hain Lifescience, Nehren, Germany). Five microlitres of the purified DNA were added to the PCR mixtures. PCR amplification was carried out using biotinylated primers specific to each assay, following the manufacturer's PCR program. After amplification, the denatured PCR products were hybridised with immobilised probes on nitrocellulose strips to detect *Mtb* wild-type (WT) sequences or mutations (MUT) associated with RIF and INH in the MTBDR*plus*, and associated with FQs (MFX and LFX), and AMK in the MTBDR*sl*. The strips were read and interpreted according to the manufacturer's instructions. A valid LPA required the presence of *Mtb* complex-specific control (TUB), conjugate controls (CC) and amplification control (AC) bands, along with the appropriate target gene locus control bands. Drug resistance was determined by the absence of one or more WT probes and the presence of a corresponding MUT probe. The absence of both WT and MUT probes also indicated resistance. Detection of all WT probes alongside any MUT probe, indicating possible heteroresistance (i.e., the presence of both susceptible and resistant strains as described by the manufacturer), was also interpreted as resistant. If all WT probes developed in the absence of MUT probes, resistance was not detected. Sample failure for the combined MTBDR*plus* and MTBDR*sl* assays was defined as any clinical sample that failed to yield interpretable results across all drug targets. This included samples where the TUB band, indicating the presence of *Mtb*, was absent, and/or where control bands for all four drug targets the gene loci were missing. Target failure was defined as a sample that failed to provide a result for a specific drug target due to the absence of *Mtb* and the lack of the control bands for the gene locus associated with that drug target.

Targeted next-generation sequencing

The end-to-end tNGS workflows and evaluation of GS and ONT have been described in detail.[18] Briefly, for GS DNA extraction from sediment samples was performed using Maxwell RSC FFPE Plus DNA kit (Promega, Madison, WI) on a Maxwell RSC DNA extraction instrument. DNA samples were processed through the Deeplex Myc-TB kit following manufacture instructions with Illumina Nextera XT DNA library preparation steps. Forty-five barcoded sample libraries and three control libraries were pooled and batch-run for 150 bp paired-end sequencing on an Illumina MiSeq instrument. The FASTQ files were uploaded into the Deeplex web-based application for an automated analysis pipeline using Deeplex Myc-TB V3.0—Extended catalogue. Mutations were reported per drug as "resistant" based on detection of drug resistance-associated variants/indels and "susceptible" when no variant/indel was detected or variant detected was not associated with drug resistance, and "uncharacterised" when detected variants could not be defined as resistant or susceptible. We defined sequencing success as GS producing sequencing results and interpreting as resistant, susceptible or uncharacterized. Failure was defined as no interpretation reported for drug targets resulting in a sample or drug target that failed to sequence or meet manufacturer analysis criteria as previously described.[18]

ONT tNGS DNA extraction was performed using Maxwell RSC PureFood Pathogen Kit (Promega, Madison, WI) for use with sediment samples on a Promega Maxwell DNA extraction instrument. Twenty-two samples with two controls were batch-processed per MinION flow cell sequencing run. The resulting FASTQ files were processed through the ONT "wf-tb-amr-3993f97d" analysis pipeline, which generated an analysis report. Drug resistance profiles were determined based on the identification of variants/indels according to the ONT interpretation. Mutations were classified as "resistant" if drug resistance-associated variants/indels were detected, and "susceptible" if no variant/indel or a variant not on their mutation list was found. Sequencing success was defined as ONT producing sequencing results and interpreting them as resistant or susceptible. Target failure was defined as a sample that failed to yield a result for a specific drug target due to the ONT pipeline reporting no amplicon targets ("undetermined").

Statistical analysis

All clinical samples that underwent testing with LPAs, GS, and ONT, and had corresponding reference data (pDST and WGS), were included in the diagnostic accuracy analysis. For each test (LPAs, GS, and ONT),

clinical sensitivity and specificity were calculated against the composite reference standard for each drug evaluated, using the epiR package (version 2.0.6). The 95% confidence intervals were computed using the Wilson score method. Because the number of invalid results differed between assays, denominators varied by assay and drug. Pair-wise differences in clinical sensitivity or specificity were tested with Fisher's exact test ($\alpha = 0.05$). For the LPAs, we additionally quantified the false-negative rate (FNR) of each individual resistance mutation relative to WGS. The FNR for each mutation was calculated as FN/(FN + TP), where FN represents the number of false negatives and TP represents the number of true positives, and FN + TP constituting the total number of positives.

To assess the agreement between resistance or susceptibility calls between LPAs and tNGS workflows, we compared binary resistance calls for each drug between LPAs vs. GS and LPAs vs. ONT. Overall concordance beyond chance was quantified with Cohen's "κ (Kappa)", accompanied by its 95% CI and a Z-test for κ (H_0: $\kappa = 0$). Because κ can be biased when class prevalence is very high or low, we also reported the Prevalence- and Bias-Adjusted κ (PABAK). Finally, we applied McNemar's paired test to the two-by-two discordance table to determine whether one assay produced disproportionately more FN or FP results than the other (H_0: discordance is symmetrical), indicating a systematic performance difference. These statistics were computed with the epiR package (v 2.0.6).

To assess diagnostic performance, we compared sample and drug target failure rates for each test. For LPAs (MTBDR*plus* and MTBDR*sl* combined), sample failure was defined as any clinical sample that did not produce an interpretable result across all drug targets (absence of TUB band indicating *Mtb* not detected, or absence of the gene locus control bands across all four drug targets). For GS, sample failure was defined as any clinical sample in which *Mtb* was not detected, resulting in no interpretation of sequence data for drug resistance. For ONT, sample failure was defined as any clinical sample in which no amplicon targets were reported by the ONT pipeline. Drug-target failure for each assay was scored when a result was missing for a specific drug although the sample was otherwise valid. Failure proportions were first described overall, and then stratified by routinely recorded proxies for bacillary load—GeneXpert MTB/RIF semi-quantitative category (High, Medium, Low, Very-low, Negative) and HIV status—and by AFB smear grade. HIV infection was retained because extensive evidence shows that people living with HIV more often produce paucibacillary sputum, a recognised risk factor for false-negative molecular results.[30]

To explore predictors of sample failure for LPAs, GS and ONT, we first fitted univariable logistic-regression models for each candidate variable. For every test we then fitted one multivariable model that always contained three *a-priori* predictors—GeneXpert MTB/RIF semi-quantitative category (High, Medium, Low, Very-low), HIV status, and RIF resistance status. GeneXpert semi-quantitative category and HIV infection both proxy bacillary load, while RIF resistance tests whether *rpoB* mutants could impair primer/probe binding. Smear microscopy grade was examined but dropped because of strong collinearity with GeneXpert category (Cramér V = 0.66). No data-driven p-value selection was applied; likelihood-ratio tests were used only to quantify the incremental contribution of individual predictors, and calibration was assessed with the Hosmer–Lemeshow test. All analyses were run in R v4.3.0 (RStudio 2023.03.0 + 386).

Interpretation plan

Diagnostic accuracy (sensitivity/specificity vs. the composite reference) was the primary outcome; pairwise differences were evaluated by Fisher's exact test and overlap of 95% CIs. Overall concordance was summarised with Cohen's κ/PABAK, and any directional bias was probed with McNemar's test. If κ indicated high agreement but McNemar was significant, we reported "high concordance with a systematic bias," giving priority to the McNemar finding because it points to clinically relevant under- or over-detection. Sample- and target-failure proportions were descriptive, and logistic models provided exploratory adjusted odds ratios for predictors of sample failure. All comparisons were prespecified; therefore, no multiplicity correction was applied, and conclusions relied on consistent patterns rather than single p-values.

Role of the funding source

The funders of the study had no role in the study design, data collection, data analysis, data interpretation, writing of the manuscript, or the decision to submit for publication.

Results

As previously reported, sputum samples from 720 participants with corresponding pDST and WGS reference data (representing 86.5% of the total enrolled population in the multicentre cross-sectional clinical study) were evaluated using GS and ONT. Here we report the LPA data and its diagnostic accuracy against the same reference standard, thereby enabling a direct comparison of LPA with GS and ONT (Fig. 1). Based on the composite reference standard (pDST and WGS), 80.1% (577/720) of participants were resistance to RIF, 72.4% (521/720) were identified as having MDR-TB, and 45.7% (315/720) were classified as having pre-XDR TB. Although 85.6% (616/720) of sputum samples were smear-positive, they varied across different GeneXpert MTB/RIF semi-quantitative categories in both RIF resistant and susceptible samples (Supplementary Fig. S2). Out of the

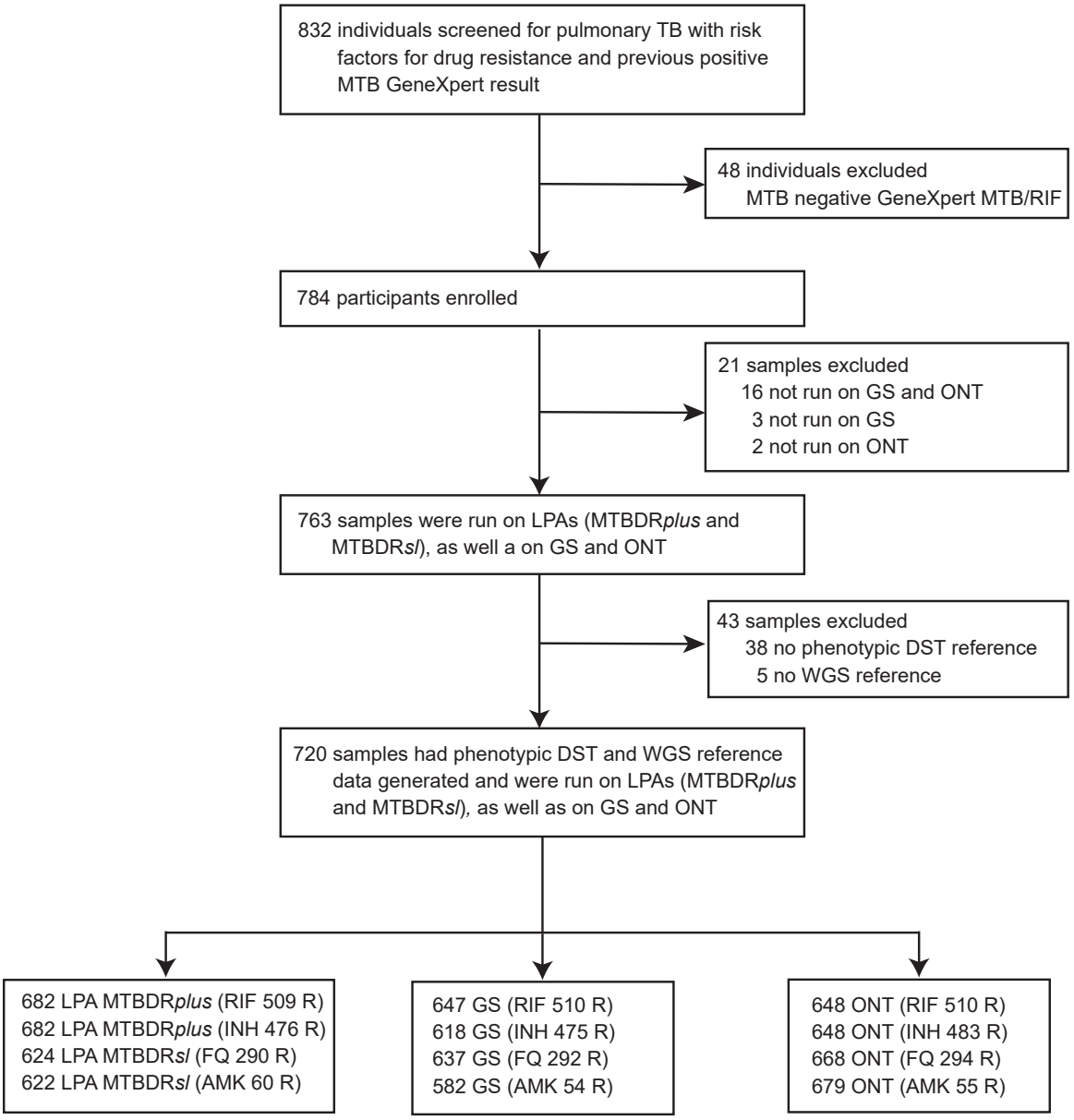


Fig. 1: **Participant enrolment and exclusions.** Risk factors and indicators for drug resistance included: 1) A positive rifampicin (RIF) resistance result by GeneXpert MTB/RIF or GeneXpert MTB/RIF Ultra; OR 2) Lack of response to TB treatment, evidenced by positive sputum smear or culture after ≥3 months of standard TB treatment; OR 3) Previous diagnosis of RIF-resistant/MDR-TB with treatment failure and positive sputum smear or culture after ≥3 months on a standard MDR-TB regimen; OR 4) Prior TB treatment of more than 1 month for a previous TB episode; OR 5) Close contact with a known drug-resistant TB case. TB = tuberculosis, MTB = *Mycobacterium tuberculosis*, GS = GenoScreen, ONT = Oxford Nanopore Technologies, LPAs = line probe assays, DST = drug susceptibility testing, WGS = whole genome sequencing, RIF = rifampicin, INH = isoniazid, FQ = fluoroquinolones, AMK = amikacin, R = resistant.

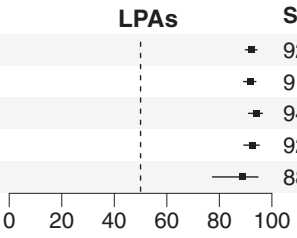
784, 64 sediments lacked a composite reference result and thus were excluded from the diagnostic accuracy analysis (Fig. 1); a best-/worst-case analysis showed that omitting them altered LPA sensitivities by ≤ 1.2 percentage points (Supplementary Methods and Table S1), confirming that their absence does not bias the accuracy estimates presented below.

Diagnostic accuracy and resistance detection yield of LPAs and tNGS compared to a composite reference standard

The LPAs (MTBDR*plus* and MTBDR*sl*) demonstrated a specificity exceeding 97% across all five drugs assessed, with sensitivity ranging from 94.3% for FQ (MFX) to 88.7% for AMK (Fig. 2a). Comparatively, as previously

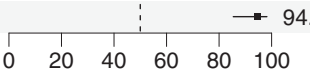
a

drug	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
RIF	682	507	2	42	131	92.3 (89.9–94.3)	98.5 (94.7–99.6)
INH	682	475	1	42	164	91.9 (89.2–93.9)	99.4 (96.6–99.9)
FQ (MFX)	624	283	7	17	317	94.3 (91.1–96.4)	97.8 (95.6–98.9)
FQ (LFX)	624	281	9	22	312	92.7 (89.3–95.2)	97.2 (94.8–98.5)
AMK	622	47	13	6	556	88.7 (77.4–94.7)	97.7 (96.1–98.7)



b

drug	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
RIF	647	510	0	14	123	97.3 (95.6–98.4)	100.0 (97.0–100.0)
INH	618	472	3	17	126	96.5 (94.5–97.8)	97.7 (93.4–99.2)
FQ (MFX)	637	287	5	8	337	97.3 (94.7–98.6)	98.5 (96.6–99.4)
FQ (LFX)	637	287	5	10	335	96.6 (93.9–98.2)	98.5 (96.6–99.4)
AMK	582	53	1	3	525	94.6 (85.4–98.2)	99.8 (98.9–100.0)



c

drug	N	TP	FP	FN	TN	Sensitivity (95% CI)	Specificity (95% CI)
RIF	648	510	0	23	115	95.7 (93.6–97.1)	100.0 (96.8–100.0)
INH	648	482	1	23	142	95.4 (93.3–96.9)	99.3 (96.1–99.9)
FQ (MFX)	668	292	2	16	358	94.8 (91.7–96.8)	99.4 (98.0–99.8)
FQ (LFX)	668	292	2	19	355	93.9 (90.7–96.1)	99.4 (98.0–99.8)
AMK	679	53	2	7	617	88.3 (77.8–94.2)	99.7 (98.8–99.9)

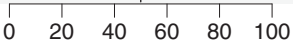


Fig. 2: **Sensitivity and specificity of LPAs, GS and ONT for detection of resistance to the five drugs assessed by LPAs compared against a composite reference standard of phenotypic drug susceptibility testing and whole-genome sequencing.** Forest plots of a) sensitivity and specificity of LPAs MTBDR*plus* and MTBDR*sl*; b) sensitivity and specificity of GenoScreen; c) sensitivity and specificity of Oxford Nanopore Technologies. Squares indicate point estimates and horizontal whiskers the 95% confidence interval of sensitivity (x-axis 0–100%). RIF = rifampicin, INH = isoniazid, FQ = fluoroquinolones, MFX = moxifloxacin, LFX = levofloxacin, AMK = amikacin, N = Number, TP = true positive, FP = false positive, FN = false negative, TN = true negative, and CI = confidence interval. For RIF, lower WHO-endorsed rifampicin critical concentration (0.5 μg/mL) was used (tNGS data are adapted from colleagues[18]).

reported, GS also exhibited >97% specificity across the same drugs, with sensitivity varying from 97.3% for RIF and FQ (MFX) to 94.6% for AMK (Fig. 2b). ONT achieved the highest specificity, exceeding 99% across all five drugs, with sensitivity ranging from 95.7% for RIF to 88.3% for AMK (Fig. 2c). The sensitivity of LPA for detecting RIF resistance was significantly lower than that of both GS and ONT (Fisher's exact test, p = 0.0003 and p = 0.0216, respectively). Similarly, LPA sensitivity for detecting INH resistance was significantly lower compared to both GS and ONT (Fisher's exact test, p = 0.0019 and p = 0.0210, respectively). The reduced sensitivity of MTBDRplus for detecting RIF and INH resistance was primarily due to (i) MUT-probe hybridisation failures or (ii) hybridisation of WT probes in the presence of resistance mutations (Supplementary Figs. S3 and S4). The sensitivity of LPA (MTBDR*sl*) for detecting FQ and AMK resistance was comparable to that of ONT (Fig. 2). Although GS achieved the highest sensitivity for detecting resistance to FQ and AMK, these differences did not reach statistical significance.

The adjusted resistance detection yield, defined as the proportion of true positives samples detected by the diagnostic test relative to the total resistance calls determined by the composite reference, was evaluated for RIF, INH, FQ, and AMK. For RIF, LPA achieved a detection yield of 87.9% (507/577), comparable to 88.4% (510/577) for both GS and ONT (Table 1). A similar pattern was observed for INH, where LPA detected 88.5% (475/537) of INH resistance, slightly lower than ONT at 89.8% (472/537), but marginally higher than GS 87.9% (472/537). For FQ, adjusted resistance detection yields ranged from 86.2% (282/

	RIF-R n (%)	INH-R n (%)	FQ-R n (%)	AMK-R n (%)
Composite reference (pDST + WGS)	577	537	327	62
GeneXpert	549 (95.1)			
pDST	532 (92.2)	536 (99.8)	313 (95.7)	60 (96.8)
WGS	566 (98.1)	511 (95.2)	309 (94.5)	56 (90.3)
LPAs	507 (87.9)	475 (88.5)	282 (86.2)	47 (75.8)
GS	510 (88.4)	472 (87.9)	287 (87.8)	53 (85.4)
ONT	510 (88.4)	482 (89.8)	292 (89.3)	53 (85.4)

Data are reported as (%), with percentages calculated relative to the total resistance calls (R) determined by the composite reference standard (pDST and WGS) for each drug, unless otherwise indicated. For GeneXpert, LPAs, GS, and ONT true resistance calls (TP) against the composite reference are reported.

Table 1: **Resistance yield for drug targets (RIF, INH, FQ, AMK) among 720 samples tested with multiple diagnostics, GeneXpert, phenotypic drug susceptibility testing (pDST), whole-genome sequencing (WGS), line probe assays (LPAs), GenoScreen (GS), and Oxford Nanopore Technologies (ONT).**

327) for LPA to 89.3% (292/327) for ONT, with GS achieving 87.8% (287/327). AMK resistance detection yields were lower across all assays, with LPA at 75.8% (47/62), which was notably lower than GS and ONT, both of which achieved 85.4% (53/62). These findings indicate that GS and ONT slightly outperformed LPAs in detecting true resistance, particularly for AMK, while maintaining comparable yields for RIF, INH, and FQ resistance.

False negative rates (FNR) of resistance mutations in LPAs compared to WGS

To further investigate the reduced sensitivity of LPA in detecting RIF and INH resistance and to assess potential of systematic differences in detecting certain mutations, we calculated the FNR for each mutation from each LPA result, using WGS as the reference method. The FNR for LPA MTBDR*plus* for detecting RIF resistance mutations compared to WGS varied substantially, ranging from 0% to 100%. For RIF resistance mutations specifically targeted by the *rpoB* MUT probes, the FNR ranged from 0% to 6.2%, aligning closely with the overall FNR for all RIF resistance mutations (7%). However, the wide variability in FNR for less frequently observed RIF mutations (0%–100%) underscores that LPA sensitivity is highly mutation-dependent (Supplementary Fig. S3). Similarly, the FNR of LPA MTBDR*plus* for detecting INH resistance mutations compared to WGS also ranged from 0% to 100%. INH mutations detected by *katG* MUT and *inhA* MUT probes alone exhibited FNRs ranging from 0% to 1.4%, closely matching the overall FNR for all INH resistance mutations (3%) (Supplementary Fig. S4). These findings suggest that systematic bias in detecting less frequently observed RIF and INH resistance mutations cannot be ruled out for the LPA MTBDR*plus*.

For the LPA MTBDR*sl*, the FNR for FQ and AMK resistance mutations compared to WGS are presented in Supplementary Figs. S5 and S6. Unlike the MTBDR*plus*, the FNR for FQ resistance mutations targeted by the *gyrA* and *gyrB* MUT probes ranged from 0% to 40% (Supplementary Fig. S5), which exceeds the overall FNR for all FQ resistance mutations (5%) This suggests that MTBDR*sl* does not exhibit systematic bias in missing less common FQ resistance mutations.

Agreement in drug resistance detection between LPA vs. GS and LPA vs. ONT

Agreement between LPAs and the two tNGS workflows was generally high: Cohen's "κ (Kappa)" ranged from 0.83 to 0.96 across all drug–assay pairs (Supplementary Table S2), and the corresponding Z-tests confirmed that every κ differed significantly from zero. However, McNemar's test revealed relevant asymmetry discordance for certain drugs. Compared with GS, LPA produced significantly more discordant results for RIF and INH (McNemar $p = 9.6 \times 10^{-6}$ and 1.6×10^{-4}, respectively, Supplementary Table S2), indicating that LPA missed resistance in a subset of clinical samples detected by tNGS. No such asymmetry was observed for FQ or AMK ($p > 0.05$). When LPAs were compared with ONT, the same pattern held for RIF and INH ($p = 0.020$ and 0.007, respectively) and extended to AMK ($p = 0.002$), while FQ calls remained symmetrical ($p = 0.49$). Thus, while κ suggests strong overall agreement, McNemar's test highlights systematic under-detection of RIF, INH and (vs. ONT) AMK resistance by LPAs, most likely due to their more limited mutation coverage.

Comparison of LPA and tNGS sample and drug target failures

Sputum samples from 720 participants were tested using LPAs and tNGS and could be evaluated for failure rate analysis (Fig. 1). For LPAs (MTBDR*plus* and MTBDR*sl*), the overall sample failure rate was low with 4.9% (35/720) of clinical sediment samples yielding no drug resistance profiles across all four drugs (Table 2). When stratified by bacterial load (as determined by the GeneXpert MTB/RIF semi-quantitative result categories), sample failure rates were highest (53.1%) in the 'very low' category and lowest (0.3%) in the 'high category. Overall, per drug target failure rates (excluding full sample failures) ranged from 0.6% to 1.4% (Table 2). FQ and AMK exhibited the highest drug target failure rates, with 1.1.% (8/720) and 14% (10/720) of clinical sediment samples failing to yield resistance profiles for FQ and AMK, respectively.

For GS, the overall sample failure rate was moderately low, with 8.6% (62/720) of clinical sediment samples failing to generate sequencing data and thus providing no resistance profiles for any of the four drugs covered by LPAs (Table 3). When stratified by bacterial load, sample failure rates for GS were highest (87.5%) in the 'very low' category and lowest (1.1%) in

	Xpert Results Category				
	High n (%)	Medium n (%)	Low n (%)	Very Low n (%)	Total n (%)
Sample failure MTBDR*plus* & MTBDR*sl*	1 (0.3)	1 (0.4)	16 (18.2)	17 (53.1)	35 (4.9)
Sample failure MTBDR*plus*	1 (0.3)	1 (0.4)	17 (19.3)	17 (53.1)	34 (4.7)
Sample failure MTBDR*sl*	10 (2.8)	26 (10.9)	30 (34.1)	22 (68.8)	88 (12.2)
RIF target failure	0 (0)	0 (0)	1 (1.1)	3 (9.4)	4 (0.6)
INH target failure	0 (0)	0 (0)	1 (1.1)	3 (9.4)	4 (0.6)
FQ target failure	0 (0)	1 (0.4)	6 (6.8)	1 (3.1)	8 (1.1)
AMK target failure	0 (0)	1 (0.4)	8 (9.1)	1 (3.1)	10 (1.4)
Total samples	362	238	88	32	720

Sample failure indicates that no results were obtained. Specifically, sample failure MTBDR*plus* and MTBDR*sl*: absence of TUB band (i.e., no *Mtb* detected), or absence of the gene locus control bands for all drug targets. Sample failure MTBDR*plus*: absence of TUB band. Sample failure MTBDR*sl*: absence of TUB band. Target failure (RIF, INH, FQ, AMK) indicates no valid result specifically due to the absence of gene locus control band for that drug target. High, Medium, Low, Very Low refer to GeneXpert's semi-quantitative bacterial load categories. N (%) represents the count (N) and proportion (%) of samples within each category.

***Table 2*: Sample and drug-target failures for the first-line (MTBDR*plus*) and second-line (MTBDR*sl*) LPAs, stratified by GeneXpert MTB/RIF semi-quantitative categories.**

the 'high' category (Table 3). Overall, per drug target failure rates (excluding full sample failures) ranged from 0% to 1.3% (Table 3). INH had the highest failure rate (1.3%), with a total 9 samples lacking INH resistance profiles.

For ONT, the overall sample failure rate was low with 5.0% (36/720) clinical sediment samples producing no sequencing data and thus no resistance profiles across all four drugs (Table 4). When sample failure was stratified by bacterial load, sample failure rates for ONT were highest (34.4%) in 'very low' and the lowest in the 'high' (1.1%) category (Table 4). Overall, per drug target failure rates (excluding full sample failures) ranged from 0.7% to 5.0%. INH and RIF had the highest failure rate, with a total of 36 samples lacking INH or RIF resistance profiles.

LPA and tNGS sample failure and their association with bacillary load and HIV status

Beyond accuracy, a molecular assay must reliably generate a result from routine specimens. Because failure most often reflects low bacillary load or inhibitors in sputum, we examined how failure rates varied by two clinical proxies for bacillary burden—HIV status and AFB smear grade—and then used multivariable modelling to quantify the independent effect of bacterial load (GeneXpert semi-quantitative results category), HIV infection and RIF resistance on assay success.

Across all assays, failure rates varied by HIV status (Table 5). Among people living with HIV, LPAs had the lowest failure rate (7.2%), followed by GS (14.4%) and ONT (19.6%); in HIV-negative specimens the ordering reversed, with ONT lowest (2.6%), then LPAs (5.3%) and GS (8.6%). As expected, the sample failure rate was higher for smear-negative than smear-positive samples across the tests (Table 5). ONT had the lowest failure rate for smear-negative samples (22.3%), followed by LPAs (33.0%), and GS (37.7%). In smear-positive samples, LPAs had the lowest failure rate (0.2%) compared with ONT (2.1%) and GS (3.7%).

Multivariable logistic regression confirmed that bacterial load, captured by GeneXpert semi-quantitative results category, was the dominant predictor for every assay. For LPAs, compared to the "high" category, "low" bacterial load was associated with OR = 0.013 (95% CI: 0.0007–0.067, $p < 0.001$) and "very low" load had OR = 0.0028 (95% CI: 0.00014–0.0156, $p < 0.001$), both indicating a markedly reduced likelihood of success. The "medium" category did not differ significantly

	Xpert Results Category				
	High n (%)	Medium n (%)	Low n (%)	Very Low n (%)	Total n (%)
Sample failure GS	4 (1.1)	9 (3.8)	21 (23.9)	28 (87.5)	62 (8.6)
RIF target failure	0 (0)	0 (0)	0 (0)	0 (0)	4 (0.6)
INH target failure	2 (0.6)	2 (0.8)	5 (5.7)	0 (0)	4 (0.6)
FQ target failure	0 (0)	1 (0.4)	1 (1.1)	0 (0)	8 (1.1)
AMK target failure	2 (0.6)	2 (0.8)	1 (1.1)	0 (0)	10 (1.4)
Total samples	362	238	88	32	720

***Table 3*: GS sample and drug target failures stratified by bacterial load (GeneXpert Results category) for only the drug targets included in the LPAs (GS data are adapted from Colman and colleagues[18]).**

	Xpert Results Category				
	High n (%)	Medium n (%)	Low n (%)	Very Low n (%)	Total n (%)
Sample failure ONT	4 (1.1)	4 (1.7)	17 (19.3)	11 (34.4)	36 (5.0)
RIF target failure	11 (3.0)	8 (3.4)	10 (11.4)	7 (21.9)	36 (5.0)
INH target failure	11 (3.0)	10 (4.2)	10 (11.4)	5 (15.6)	36 (5.0)
FQ target failure	0 (0)	3 (1.3)	8 (9.1)	5 (15.6)	16 (2.2)
AMK target failure	0 (0)	1 (0.4)	0 (0)	4 (12.5)	5 (0.7)
Total samples	362	238	88	32	720

Table 4: **ONT sample and drug target failures stratified by bacterial load (GeneXpert Results category) for only the drug targets included in the LPAs (ONT data are adapted from Colman and colleagues[18]).**

from "high" (p = 0.792). Neither RIF resistance (p = 0.974) nor HIV status (p = 0.219) significantly impacted LPA success. The Hosmer–Lemeshow test indicated an adequate fit (p = 0.143).

A similar pattern emerged for GS, although the "medium" category showed a borderline significant effect. Compared to the "high" bacterial load group, "low" had OR = 0.029 (95% CI: 0.0065–0.088, p < 0.001), "medium" had OR = 0.25 (95% CI: 0.055–0.89, p = 0.044), and "very low" had OR = 0.0010 (95% CI: 0.00015–0.0048, p < 0.001). Neither RIF resistance (p = 0.229) nor HIV status (p = 0.617) significantly affected GS success, and model fit was good (p = 0.886).

For ONT, both bacterial load and HIV infection influenced outcomes. Compared to "high" load, "low" had OR = 0.064 (95% CI: 0.014–0.208, p < 0.001) and "very low" had OR = 0.023 (95% CI: 0.0046–0.091, p < 0.001), indicating markedly lower chances of success. HIV positivity further reduced success (OR = 0.157, 95% CI: 0.067–0.358, p < 0.001), whereas RIF resistance had no effect. Model fit remained acceptable (p = 0.831).

	MTBDR*plus* + MTBDR*sl* Sample failures n (%)	GenoScreen Sample failures n (%)	ONT Sample failures n (%)	Total n
HIV status				
HIV negative	28 (5.3)	46 (8.6)	14 (2.6)	533
HIV positive	7 (7.2)	14 (14.4)	19 (19.6)	97
HIV unknown	0 (0)	2 (2.2)	3 (3.3)	90
Total	35 (4.9)	62 (8.6)	36 (5.0)	720
Smear status				
Smear negative	34 (33.0)	39 (37.7)	23 (22.3)	103
Smear positive	1 (0.2)	23 (3.7)	13 (2.1)	616
Smear unknown	0 (0)	0 (0)	0 (0)	1
Total	35 (4.9)	62 (8.6)	36 (5.0)	720

Sample failure for LPAs was defined as any clinical sample that did not produce a call (resistance non detected, resistance detected, resistance inferred) across all drug targets interrogated by LPAs (MTBDR*plus* and MTBDR*sl*) (tNGS data are adapted from colleagues[18]).

Table 5: **Comparison of sample failure rates stratified by HIV and AFB smear status for LPAs, GS and ONT.**

Discussion

In this multicentre, cross-sectional study, we evaluated and compared the diagnostic performance of LPAs and two end-to-end tNGS workflows, GenoScreen (GS), and Oxford Nanopore Technologies (ONT), for detecting drug resistant tuberculosis in high-burden settings using the same clinical sediment samples. Our findings revealed that LPA (MTBDR*plus*) exhibited significantly lower diagnostic accuracy for detecting resistance to RIF and INH compared to tNGS workflows when evaluated against a composite reference standard. Specifically, LPA demonstrated sensitivity of approximately 92% for RIF and INH resistance detection, while GS and ONT achieved higher sensitivities, exceeding 96% and 95%, respectively. The lower sensitivities of LPA observed in this study aligns to previous reports, which have also noted variable performance, particularly lower sensitivity for INH resistance detection compared to RIF.[31,32]

Given the WHO's recommendations for the use of new and repurposed drugs (BPaLM, BPaL)[11] that are currently not detectable by LPA, it is essential to rule out INH and FQ resistance. Many patients started on the short MDR-RR TB regimen are diagnosed via Xpert MTB/RIF, which does not detect INH or FQ resistance. Thus, the additional information provided by LPAs (MTBDR*plus* and MTBDR*sl*) on INH and FQ resistance is crucial for clinicians. However, the marginally lower sensitivity of LPA for these drugs means that some cases will be missed, highlighting a potential adjunctive role for tNGS in the early management of MDR-RR TB. Our data suggest that GS, with a sensitivity of 97% for FQ resistance detection, outperforms LPA (MTBDR*sl*) and ONT, both of which demonstrated sensitivities below 95%. High sensitivity for FQ resistance detection, may serve as a valuable addition to the diagnostic toolkit for ruling out FQ resistance, ensuring minimal missed cases and supporting optimal treatment regimens for MDR-RR and pre-XDR-TB alongside existing diagnostic approaches like MTBDR*sl* LPA.

The observed diagnostic accuracy of the two end-to-end tNGS workflows, demonstrates consistently high sensitivity for RIF and INH resistance compared to

LPAs on the same sediment samples. This finding underscores tNGS utility in diagnostic algorithms, particularly in settings with diverse and complex resistance profiles, where precise and comprehensive drug resistance profiling is vital for guiding effective treatment. Further gains in sensitivity and specificity are likely as bioinformatics pipelines incorporate an expanding catalogue of resistance-associated mutations.[18,27] Future releases of updated sequencers and tNGS kits, such as ONT's GridION-Q and updated flow-cell chemistries, which promises higher read depth and may reduce sample-failure rates, however sequencing read accuracy will need to be evaluated.

Our findings on sample success rates have important implications for the implementation of diagnostic assays in high-burden tuberculosis settings. Both LPAs and ONT demonstrated high and comparable overall sample success rates, with 685 (95.1%) and 684 (95.0%) out of 720 clinical sediment samples, respectively, producing at least partial drug resistance profiles. In contrast, GS had a slightly lower overall success rate, with 658 (91.4%) samples producing at least partial drug resistance profiles. Sample failure rates for all three assays, were largely driven estimated bacillary burden in the sediment. GeneXpert MTB/RIF semi-quantitative category was the dominant predictor, with yields dropping sharply from "low" to "very low," and, for GS, already from "medium" to "low". For ONT alone, HIV positivity was also a significant predictor of sample failure. These findings align with previous evidence highlighting the critical role of bacterial load in determining diagnostic success.[18,33] Furthermore, the differences in sample failure by HIV status likely reflects technical differences in DNA extraction and/or library preparation efficiency, which affected samples with low bacterial load and ONT workflow more than others in this study. These findings emphasize the importance of optimising steps like DNA extraction, especially for low bacterial load samples, to enhance assay performance and ensure their applicability across diverse clinical settings as with any molecular assay. Practical optimisation steps—ranging from low-cost bead-beating or lysozyme/SDS pre-lysis to higher-end host-DNA-depletion or hybrid-capture enrichment—can be layered onto existing workflows to increase mycobacterial DNA yield and thus sequencing success. Although, it is important to note that both tNGS workflows evaluated had specified optimisations in place for the DNA extraction section of their protocols.

Our findings highlight important differences in assay performance at sample- and drug-target levels. Although the overall sample-level failure rate for LPAs (MTBDR*plus* and MTBDR*sl* combined) was relatively low (4.9%), MTBDR*sl* exhibited higher per-target failure for FQ (1.1%) and for AMK (1.4%) when excluding full sample failures. When sample-level MTBDR*sl* failures are factored in, however, the total number of clinical samples lacking FQ and AMK resistance profiles increases substantially, to 96 samples (13.3%) for FQ and 98 samples (13.6%) for AMK (Table 2). By contrast, GS had very low per-target failures for FQ (0.3%) and AMK (0.7%) (excluding full sample failures), whereas ONT showed a higher FQ failure rate (2.2%) but a similar AMK rate (0.7%). Despite their strong overall performance, GS and ONT recorded their highest failure rates with INH (1.3% for GS and 5% for ONT), surpassing the 0.6% observed for MTBDR*plus*. Because such target-specific drop-outs typically reflect technical factors (e.g., primer mismatches, secondary structure, or depth thresholds) rather than patient-level covariates, they were summarised descriptively rather than modelled. Collectively, these observations underscore the need to consider both sample-level and individual drug-target performance and diagnostic accuracy when selecting or interpreting diagnostic assays for MDR-TB and pre-XDR TB regimens.

Additionally, our results indicate that LPA (MTBDR*plus*) systematically miss less common drug resistance mutations associated with RIF and INH resistance included in the assays, contributing to undetected resistance. This finding is consistent with a previous meta-analysis.[7] The variability in LPA false negative rate per RIF and INH resistance mutations—ranging from 0% to 100%—underscores that the LPA's sensitives are highly mutation-dependent. By calculating the probabilities and likelihood ratio to quantify the relative risk of FN results for probe types (MUT vs. WT probes), we found that FN results for RIF resistance detection using LPA were 48 times more likely when mutations in the *Mtb* sample are associated with WT probes compared to MUT probes. Similarly, for INH resistance, the relative risk of FN results was over 100 times higher for WT-associated mutations. These findings indicate that LPA (MTBDR*plus*) systematically misses less frequent, resistance-conferring mutations associated with WT probes due to its reliance on MUT probes targeting common mutations. This gap contributes to undetected RIF and INH resistance, raising concerns about diagnostic algorithms that depend on MTBDR*plus* for ruling out resistance of RIF and INH, particularly in settings with diverse drug resistance patterns and mutation variability.

Furthermore, our findings highlight a critical limitation of LPA MTBDR*sl* in detecting FQ resistance mutations, with FNRs for FQ resistance mutations targeted by *gyrA* and *gyrB* MUT probes ranging from 0% to 40%, exceeding the overall FNR for all FQ resistance mutations (5%). While no obvious systematic bias was observed for MTBDR*sl* in missing less common FQ resistance mutations, the observed higher FNR for some mutations combined with the high FQ target failure rates for FQ, may compromise the assay's ability to reliably identify FQ resistance in clinical settings. Given the essential role of FQ in the shorter

MDR-RR TB regime, these findings further emphasize the need for diagnostic tools with consistent and high sensitivity across all known resistance mutations.

Our study has several limitations. First, the high proportion of the sputum samples being smear-positive may not fully represent populations with higher prevalence of smear-negative or paucibacillary TB, which are more challenging to diagnose and treat. A key technical limitation observed was the impact of low bacterial load on LPAs and sequencing success, which present significant challenges in settings with a high prevalence of paucibacillary TB samples. Second, this study was designed for diagnostic-accuracy endpoints and did not collect primary data on turnaround time, staffing, or per-sample cost; these operational factors will ultimately govern adoption, particularly in resource-constrained laboratories. Nevertheless, the recent cost-effectiveness evaluation that use our accuracy inputs on tNGS (multi-country modelling analysis by Shrestha and colleagues[34]), projects that commercial tNGS workflows would already be cost-saving or cost-effective in certain settings. Targeted implementation studies that evaluate costs, ease-of-use, and local supply-chain robustness will be crucial to define the best-use cases for the LPA vs. tNGS and ensure optimum impact in improving outcomes for patients with DR-TB.

Contributors

GC, REC, MS and TCR conceptualised the study. REC, AS, SU, and TCR acquired the funding. GC and MS conducted the formal analysis. ADLR, CR, NT, NM, and SVO acquired the data. GC, REC, MS, AS, SU, and TCR had full access to the data in the study. GC, REC and MS accessed and verified the data. GS drafted the original manuscript. All authors read and approved the final version of the manuscript.

Data sharing statement

The relevant data are presented in the manuscript. The tNGS sequence data are available in the NCBI Sequence Read Archive under the accession number PRJNA1160005. Additional LPA and tNGS data can be accessed via the Dryad Digital Repository at https://doi.org/10.5061/dryad.gqnk98t0f and https://doi.org/10.5061/dryad.qjq2bvqs6 respectively. The R code used to perform the statistical analyses in this study is available at GitHub https://github.com/giocarpi/TB_drug_resistance_LPA_tNGS.

Declaration of interests

TCR, MS, REC received salary support from FIND through a service contact to UC San Diego. TCR and REC Received grant funding from NIH to develop and evaluate a tNGS solution for drug resistant TB (R01AI176401). TCR and REC are co-inventors on a patent associated with the processing of TB sequencing data (European Patent Application No. 14840432.0 & USSN 14/912,918). Both TCR and REC have transferred all rights and present and future interest in and rights to royalties from this patent to UC San Diego and Translational Genomics Research Institute respectively. TCR is a co-founder, board member and unpaid shareholder of Verus Diagnostics Inc, a company that was founded with the intent of developing diagnostic assays. Verus Diagnostics is not pursuing any drug resistant TB diagnostics nor any diagnostics related to the technology or approaches discussed or mentioned in this manuscript. Verus Diagnostics was not involved in any way with data collection, analysis or publication of the results of this manuscript. TCR has not received any financial support from Verus Diagnostics. CR has received honoraria payments from Becton Dickinson and she is on the scientific advisory board for Cepheid and bioMérieux. All other authors declare no competing interests.

Acknowledgements

Support for this project was provided through funding from Unitaid (2019-32-FIND MDR). The views expressed by the authors do not necessarily reflect the views of the funding agency. We thank the study participants and their families for generously volunteering to participate in this study and the study sites for their time and effort in conducting the study and assisting with the analysis of the operational data.

Appendix A. Supplementary data

Supplementary data related to this article can be found at https://doi.org/10.1016/j.ebiom.2025.105875.

References

1 World Health Organization. *Global tuberculosis report 2023*. 2023. Geneva.
2 World Health Organization. *Global tuberculosis report 2024*. Geneva: World Health Organization; 2024. Contract No.: Licence: CC BY-NC-SA 3.0 IGO.
3 World Health Organization. *Global tuberculosis report 2022*. 2022. Geneva.
4 World Health Organization. WHO consolidated guidelines on tuberculosis. In: *Module 3: Diagnosis Rapid Diagnostic for tuberculosis detection*. 2021.
5 Global Laboratory Initiative. *Line probe assays for detection of drug-resistant tuberculosis: interpretation and reporting manual for laboratory staff and Clinicians*. 2022. Contract No.: Licence: CC BY-NC-SA 3.0 IGO.
6 Seifert M, Georghiou SB, Catanzaro D, et al. MTBDRplus and MTBDRsl assays: absence of wild-type probe hybridization and implications for detection of drug-resistant tuberculosis. *J Clin Microbiol*. 2016;54(4):912–918.
7 Nathavitharana RR, Cudahy PG, Schumacher SG, Steingart KR, Pai M, Denkinger CM. Accuracy of line probe assays for the diagnosis of pulmonary and multidrug-resistant tuberculosis: a systematic review and meta-analysis. *Eur Respir J*. 2017;49(1).
8 Schön T, Miotto P, Köser CU, Viveiros M, Böttger E, Cambau E. Mycobacterium tuberculosis drug-resistance testing: challenges, recent developments and perspectives. *Clin Microbiol Infect*. 2017;23(3):154–160.
9 Ajileye A, Alvarez N, Merker M, et al. Some synonymous and nonsynonymous gyrA mutations in Mycobacterium tuberculosis lead to systematic false-positive fluoroquinolone resistance results with the hain GenoType MTBDRsl assays. *Antimicrob Agents Chemother*. 2017;61(4).
10 Fitzgibbon MM, Roycroft E, Sheehan G, et al. False detection of rifampicin resistance using Xpert((R)) MTB/RIF ultra assay due to an A451V mutation in Mycobacterium tuberculosis. *JAC Antimicrob Resist*. 2021;3(3):dlab101.
11 World Health Organization. *Rapid communication: key changes to the treatment of drug-resistant tuberculosis (WHO/UCN/TB/2022.2)*. 2022. Geneva.
12 Dolinger DL, Colman RE, Engelthaler DM, Rodwell TC. Next-generation sequencing-based user-friendly platforms for drug-resistant tuberculosis diagnosis: a promise for the near future. *Int J Mycobacteriol*. 2016;5(Suppl 1):S27–S28.
13 Leung KS, Tam KK, Ng TT, et al. Clinical utility of target amplicon sequencing test for rapid diagnosis of drug-resistant Mycobacterium tuberculosis from respiratory specimens. *Front Microbiol*. 2022;13:974428.
14 Kambli P, Ajbani K, Kazi M, et al. Targeted next generation sequencing directly from sputum for comprehensive genetic information on drug resistant Mycobacterium tuberculosis. *Tuberculosis (Edinb)*. 2021;127:102051.
15 Colman RE, Mace A, Seifert M, et al. Whole-genome and targeted sequencing of drug-resistant Mycobacterium tuberculosis on the iSeq100 and MiSeq: a performance, ease-of-use, and cost evaluation. *PLoS Med*. 2019;16(4):e1002794.
16 Kandler JL, Mercante AD, Dalton TL, et al. Validation of novel Mycobacterium tuberculosis isoniazid resistance mutations not detectable by common molecular tests. *Antimicrob Agents Chemother*. 2018;62(10):e00974-18.

17 Schwab TC, Perrig L, Goller PC, et al. Targeted next-generation sequencing to diagnose drug-resistant tuberculosis: a systematic review and meta-analysis. *Lancet Infect Dis*. 2024;24(10):1162–1176.
18 Colman RE, Seifert M, De la Rossa A, et al. Evaluating culture-free targeted next-generation sequencing for diagnosing drug-resistant tuberculosis: a multicentre clinical study of two end-to-end commercial workflows. *Lancet Infect Dis*. 2024;25(3):325–334.
19 Smith CHT, Shea J, Modestil H, et al. Assessing nanopore sequencing for clinical diagnostics: a comparison of next-generation sequencing (NGS) methods for mycobacterium tuberculosis. *J Clin Microbiol*. 2020;59(1).
20 Tafess K, Ng TTL, Lao HY, et al. Targeted-sequencing workflows for comprehensive drug resistance profiling of Mycobacterium tuberculosis cultures using two commercial sequencing platforms: Comparison of analytical and diagnostic performance, turnaround time, and cost. *Clin Chem*. 2020;66(6):809–820.
21 Murphy SG, Smith C, Lapierre P, et al. Direct detection of drug-resistant Mycobacterium tuberculosis using targeted next generation sequencing. *Front Public Health*. 2023;11:1206056.
22 Jouet A, Gaudin C, Badalato N, et al. Deep amplicon sequencing for culture-free prediction of susceptibility or resistance to 13 anti-tuberculous drugs. *Eur Respir J*. 2021;57(3).
23 Bonnet I, Enouf V, Morel F, et al. A comprehensive evaluation of GeneLEAD VIII DNA platform combined to deeplex Myc-TB((R)) assay to detect in 8 days drug resistance to 13 antituberculous drugs and transmission of Mycobacterium tuberculosis complex directly from clinical samples. *Front Cell Infect Microbiol*. 2021;11:707244.
24 World Health Organization. *Use of targeted next-generation sequencing to detect drug-resistant tuberculosis: rapid communication, July 2023*. 2023.
25 World Health Organization. *Technical report on critical concentrations for drug susceptibility testing of isoniazid and the rifamycins (rifampicin, rifabutin and rifapentine)*. 2021. Geneva.
26 World Health Organization. *Catalogue of mutations in mycobacterium tuberculosis complex and their association with drug resistance*. Geneva: World Health Organization; 2021. Contract No.: Licence: CC BY-NC-SA 3.0 IGO.
27 Walker TM, Miotto P, Koser CU, et al. The 2021 WHO catalogue of Mycobacterium tuberculosis complex mutations associated with drug resistance: a genotypic analysis. *Lancet Microbe*. 2022;3(4):e265–e273.
28 Hain Lifescience. *GenoType MTBDRplus VER 2.0. Instructions for use. IFU-304A-09*. 2023.
29 Hain Lifescience. *GenoType MTBDRsl VER 2.0. Instructions for use. IFU-317A-04*. 2023.
30 Hanrahan CF, Theron G, Bassett J, et al. Xpert MTB/RIF as a measure of sputum bacillary burden. Variation by HIV status and immunosuppression. *Am J Respir Crit Care Med*. 2014;189(11):1426–1434.
31 Lanzas F, Ioerger TR, Shah H, Acosta W, Karakousis PC. First evaluation of GenoType MTBDRplus 2.0 performed directly on respiratory specimens in central America. *J Clin Microbiol*. 2016;54(10):2498–2502.
32 Penn-Nicholson A, Georghiou SB, Ciobanu N, et al. Detection of isoniazid, fluoroquinolone, ethionamide, amikacin, kanamycin, and capreomycin resistance by the Xpert MTB/XDR assay: a cross-sectional multicentre diagnostic accuracy study. *Lancet Infect Dis*. 2022;22(2):242–249.
33 Mansoor H, Hirani N, Chavan V, et al. Clinical utility of target-based next-generation sequencing for drug-resistant TB. *Int J Tuberc Lung Dis*. 2023;27(1):41–48.
34 Shrestha S, Addae A, Miller C, Ismail N, Zwerling A. Cost-effectiveness of targeted next-generation sequencing (tNGS) for detection of tuberculosis drug resistance in India, South Africa and Georgia: a modeling analysis. *EClinicalMedicine*. 2025;79:103003.