PNAS

RESEARCH ARTICLE | BIOPHYSICS AND COMPUTATIONAL BIOLOGY

Rapid, antibiotic incubation- free determination of tuberculosis drug resistance using machine learning and Raman spectroscopy

Babatunde Ogunlade[a,1,2], Loza F. Tadesse[b,c,d,e,1], Hongquan Li[f], Nhat Vu[g], Niaz Banaei[h], Amy K. Barczak[d,i,j], Amr A. E. Saleh[a,k], Manu Prakash[b], and Jennifer A. Dionne[a,l,2]

Affiliations are included on p.7.

Edited by Catherine Murphy, University of Illinois at Urbana- Champaign, Urbana, IL; received September 8, 2023; accepted April 2, 2024

Tuberculosis (TB) is the world's deadliest infectious disease, with over 1.5 million deaths and 10 million new cases reported anually. The causative organism *Mycobacterium tuberculosis* (Mtb) can take nearly 40 d to culture, a required step to determine the pathogen's antibiotic susceptibility. Both rapid identification and rapid antibiotic susceptibility testing of Mtb are essential for effective patient treatment and combating antimicrobial resistance. Here, we demonstrate a rapid, culture- free, and antibiotic incubation- free drug susceptibility test for TB using Raman spectroscopy and machine learning. We collect few- to- single- cell Raman spectra from over 25,000 cells of the Mtb complex strain Bacillus Calmette-Guérin (BCG) resistant to one of the four mainstay anti -TB drugs, isoniazid, rifampicin, moxifloxacin, and amikacin, as well as a pan-susceptible wildtype strain. By training a neural network on this data, we classify the antibiotic resistance profile of each strain, both on dried samples and on patient sputum samples. On dried samples, we achieve >98% resistant versus susceptible classification accuracy across all five BCG strains. In patient sputum samples, we achieve ~79% average classification accuracy. We develop a feature recognition algorithm in order to verify that our machine learning model is using biologically relevant spectral features to assess the resistance profiles of our mycobacterial strains. Finally, we demonstrate how this approach can be deployed in resource-limited settings by developing a low- cost, portable Raman microscope that costs <$5,000. We show how this instrument and our machine learning model enable combined microscopy and spectroscopy for accurate few-to - single- cell drug susceptibility testing of BCG.

Raman spectroscopy | machine learning | tuberculosis | antibiotic susceptibility | infectious disease

The discovery of antibiotics in the early 20th century marked a turning point in our defense against tuberculosis (TB)—one of the deadliest infectious diseases known to humans. At that time, TB had become curable and was considered to be on a path toward elimination. However, by the end of the 20th century, TB re- emerged as a leading cause of death globally, in part due to challenges in diagnosis and its evolved resistance to antibiotics. While growth in liquid culture is considered the gold standard for organism identification and determination of antibiotic susceptibility, the causative organism, *Mycobacterium tuberculosis* (Mtb), can take up to 40 d to culture. As antibiotic susceptibility testing requires pathogen growth, definitive determination of Mtb antibiotic resistance using traditional antibiotic susceptibility testing (AST) can take additional weeks. In the delay between diagnosis and AST results, many patients are treated with only partially active antibiotic regimens, giving rise to resistant strains. In 2021 alone, more than half a million cases of multidrug- resistant TB infections were reported (1).

To mitigate the rise of antimicrobial resistance, a key strategy of the World Health Organization (WHO) is to improve surveillance of antibiotic-resistant infections and to promote the appropriate use of quality medicines (2). To address WHO guidelines, several culture-free AST approaches are being developed. For example, nucleic acid amplification tests and PCR do not require culturing and are highly sensitive and specific. However, these tests are limited to identifying resistance in cases in which a defined number of known genomic mutations confer resistance [e.g., the rpoB gene for resistance to rifampicin (3)]. In general, the appearance of genetic mutations in previously identified genes of interest has not been consistently linked to resistance to these antibiotics. Moreover, known genetic elements or mutations typically constitute only a fraction of the genetic basis for clinically relevant resistance (4). These assays also require multiple reagents, can suffer from errors with each thermal cycle, and cannot distinguish between live and dead bacteria, so they cannot be used to monitor treatment efficacy. Furthermore, these assays cost well above the

Significance

Tuberculosis is preventable and curable, yet it is the world's deadliest infectious disease. This is in part due to the emergence of drug- resistant tuberculosis. Timely, accurate drug susceptibility testing of *Mycobacterium tuberculosis* is critical for effective patient treatment and the

[1] B.O., and L.F.T. contributed equally to this work.

[2] To whom correspondence may be addressed. Email: bogun@stanford.edu or jdionne@stanford.edu.

This article contains supporting information online at https://www.pnas.org/lookup/suppl/doi:10.1073/pnas. 2315670121/-/ DCSupplemental.

Published June 11, 2024.

prevention of community spread, as inappropriate usage of anti-TB drugs may delay treatment progress and generate acquired resistance. Here, we develop a fast antibiotic susceptibility test for tuberculosis by combining Raman spectroscopy and machine learning. Using this methodology, we can assess the resistance of antibiotic- resistant mycobacteria strains at clinically relevant speeds and accuracies. Our findings provide a foundation for rapid, low- cost, point- of- care mycobacterial drug susceptibility testing for diagnostic and surveillance applications, a major step in the fight against antimicrobial resistance.

Competing interest statement: H.L. and M.P. are shareholders in Cephla Inc.

This article is a PNAS Direct Submission.

targeted $5,000 in capital cost specified by the WHO for next- generation TB drug-susceptibility testing tools (5). Complementing PCR, loop- mediated isothermal amplification (LAMP) assays eliminate the need for thermal cycling. However, these LAMP assays cannot detect mutations in resistance-associated genes because of their inability to resolve single- nucleotide differences (6). Rapid lateral flow point- of- care test antigen/antibody detection methods provide simpler and faster alternatives to LAMP and PCR but have poor sensitivity and specificity. There is also a significant delay in the appearance of target antigens in sputum or antibodies in the bloodstream, and these antigens do not necessarily indicate Mtb drug susceptibility. Therefore, current AST methods do not collectively provide the speed, sensitivity, and specificity needed in one system to meet WHO's goals for drug- resistant TB eradication.

Raman spectroscopy has the potential to identify the antibiotic resistance profiles of bacteria at the few-to - single- cell level using acquisition times on the order of seconds. Raman spectroscopy utilizes inelastic light scattering to probe the vibrational modes of a sample as a fingerprint. Different bacterial phenotypes are characterized by unique biomolecular compositions, leading to subtle differences in their corresponding Raman spectra. The Raman spectra of biological macromolecules typically reside between 500 and 1,900 cm^{-1}, with lipids, proteins, and nucleic acids exhibiting fingerprints between 1,000 and 1,700 cm^{-1}, 1,200 and 1,660 cm^{-1}, and 600 and 1,100 cm^{-1}, respectively (7). As such, these unique Raman spectral signatures can be used for accurate cellular detection, identification, and antibiotic susceptibility testing. Using Raman spectroscopy and machine learning (ML) based spectral analysis, we have previously shown that genetically engineered antibiotic-resistant and susceptible *Staphylococcus aureus* strains can be classified with ~89% accuracy (7). Similarly, we and others have used machine learning- assisted Raman to identify over 31 bacterial species and strains (8–11), including pathogens in liquid solvents (12) and in urine and blood (13–15); to identify substrains of *Escherichia coli* (16, 17), to identify various respiratory viruses (18), to identify various cellular metabolites and secretomes (19), to assess the effects of antibiotics on resistant and susceptible bacteria (20, 21), and to determine *S. aureus* antibiotic susceptibility without antibiotic incubation (22). Despite these pioneering studies, the robust determination of antibiotic susceptibility of TB, and methods translation to TB-endemic regions, remains an open challenge. Previous studies have demonstrated the use of Raman and machine learning for mycobacterial detection and antibiotic resistance profiling with high classification accuracies (23, 24). However, the machine learning models trained in these studies behave like black boxes, lacking transparency in their decision- making and interpretability and raising questions concerning the spectral features being used to achieve such high classification accuracies. These questions are especially salient when spectra are collected in complex physiological media such as sputum, which has significant spectral overlap with the vibrational information contained in Mtb.

Here, we demonstrate a rapid, culture-free, and antibiotic incubation- free antibiotic susceptibility test for TB, based on the integration of few-to -single -cell Raman spectroscopy and machine learning. In particular, we systematically classify the antibiotic resistance profiles of five isogenic Bacillus Calmette–Guérin (BCG) strains using a convolutional neural network. We collect Raman spectra from over 25,000 individual cells which are controllably engineered to possess resistance to one of the four mainstay anti- TB drugs (first- line drugs isoniazid and rifampicin and second-line drugs moxifloxacin and amikacin) and a pansusceptible wild-type strain. Using only 15across the entire spectral range (740 to 1,802 cm-s integration times and Raman features from $_{-1}$), we achieve

>98% resistant versus susceptible classification accuracy across all five BCG strains. In addition, we show that various genetic mutants can be accurately classified according to their antibiotic resistance. Using our feature recognition algorithm, we also identify the subsets of wavenumbers which most strongly influence antibiotic resistance classification and correlate them with the Raman signatures of well- known biomolecules and functional groups, verifying that our machine learning model is using biologically relevant peaks for classification. Our algorithm shows that antibiotic resistance behavior is primarily reflected in certain vibrational modes—particularly those of mycolic acid—that underly the molecular-level antibiotic response of the bacteria. This algorithm also shows that targeting the entire spectral range of scientific- grade Raman instruments is not necessary for TB AST. To show how our approach can be robustly deployed in resource-limited settings, we develop a low- cost portable Raman setup for point-of -care applications. Our instrument costs <$5,000, and with inclusion of low- cost plasmonic nanoantennas for surface- enhanced Raman spectroscopy (SERS), achieves an average accuracy of 89% across all five BCG strains. In dried patient sputum samples, our instrument achieves ~79% classification accuracy—providing a foundation for rapid, low- cost, and portable AST testing in resource-limited regions.

We performed our antibiotic susceptibility test on five isogenic BCG strains: those resistant to the first- line TB antibiotics isoniazid and rifampicin; those resistant to the second- line TB antibiotics moxifloxacin and amikacin; and also on a fully antibiotic-susceptible control strain (hereby described as wildtype). Fig. 1*C* shows the minimum inhibitory concentrations (MICs) of the four anti-TB antibiotics used on our five BCG strains, confirming each resistant strain's resistance to a single antibiotic as well as the full antibiotic susceptibility of the wild- type strain. As seen in Fig. 1 *D*–*H*, scanning electron micrographs show no distinct morphological difference among the resistant strains.

Results

Few-to -single -cell Raman Spectroscopy of BCG for Antibiotic Incubation-free Determination of Drug Resistance. As summarized in Fig. 1, we deposit mycobacterial cells onto gold-coated glass substrates and collect few-to -single - cell Raman spectra (Fig. 1*A*). For our studies, we controllably engineer antibiotic-resistant Mtb models using BCG, a live-attenuated form of *Mycobacterium Bovis* as a model organism. While lacking central virulence determ inants, BCG is nearly identical to Mtb in its core bacterial funct ions, including DNA replication, RNA polymerase, and cell wall functions. We use these models to controllably engineer single- antibiotic resistance for robust profiling and to enable safe handling of resistant samples in our BSL- 2 facilities for the proof- of-concept studies performed here. Individual Raman spectra are used as an input into our convolutional neural network; we employed a 10-fold cross - validation scheme to determine the unbiased model performance. The output of this algorithm is a series of probability scores for antibiotic resistance (Fig. 1*B*). The class with the highest probability score is chosen as the predicted class.

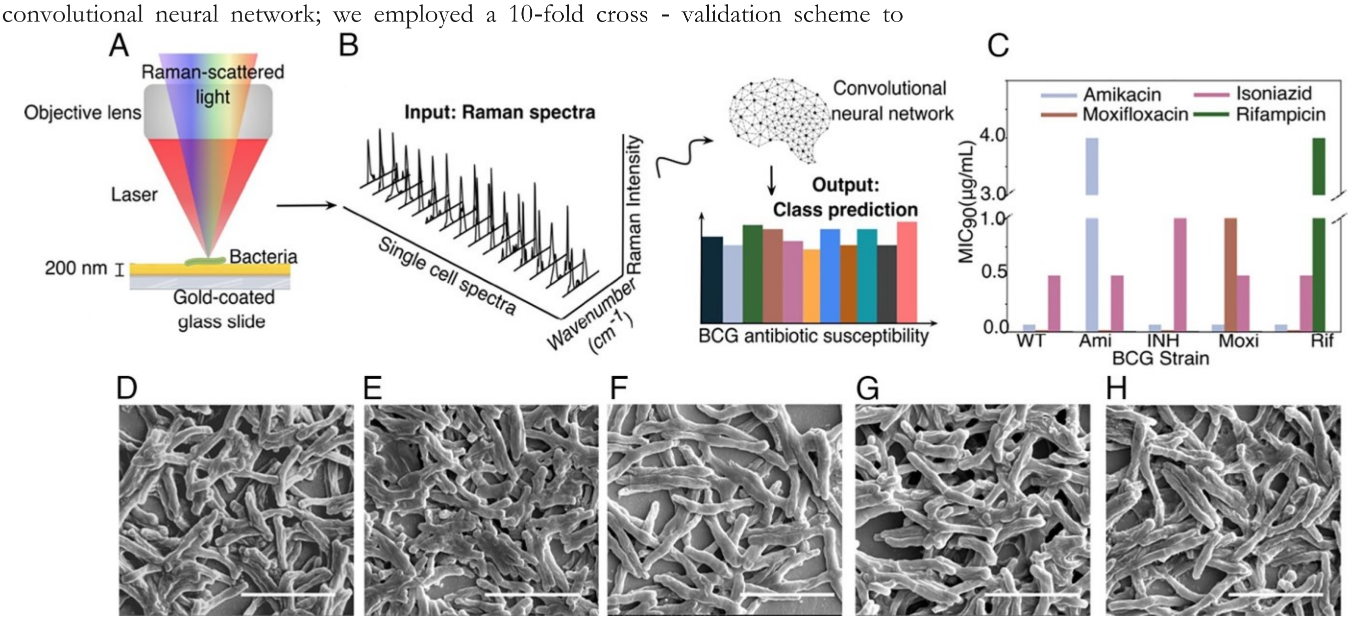
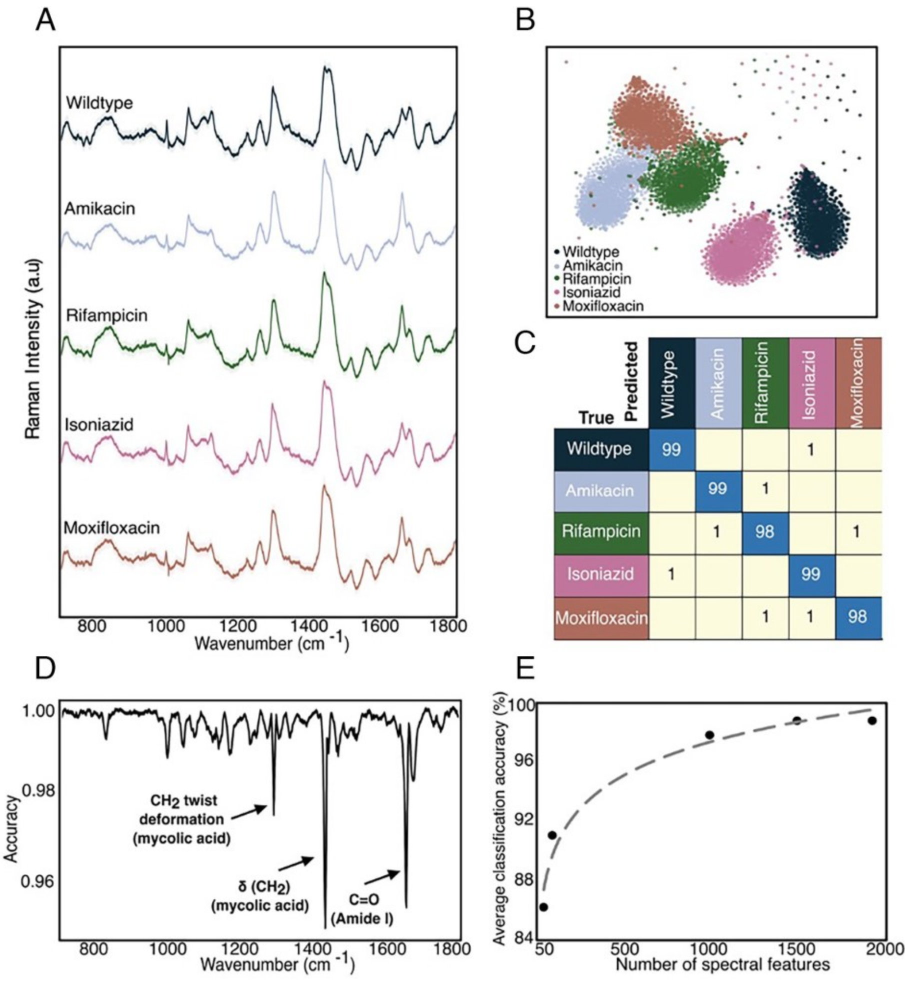


Fig. 1. Experimental setup and example bacterial Raman spectra with antibiotic resistance profiles. (*A*) Diffraction- limited spot of a 633 nm incident laser collecting a spectrum from roughly a single bacterial cell. (*B*) Bacterial spectra are used as input for the neural network to perform a classification task that outputs probability scores for each bacterial strain's antibiotic resistance. (*C*) Minimum inhibitory concentrations (MICs) for the five distinct strains tested across the four main antibiotics using standard serial two-fold dilution. See *SI Appendix*, Table S1 for details. (*D*–*H*) Scanning electron microscopy (SEM) image of a monolayer of wildtype, amikacin- resistant, rifampicin- resistant, isoniazid- resistant, and moxifloxacin- resistant BCG. (Scale bar, 4 µm.)

As shown in Fig. 2*A*, the average Raman spectra of the five distinct BCG strains (~1,700 spectra per strain) look similar, with major spectral peaks preserved across the strains. However, there are some noticeable differences in the ~1,000, ~1,300, ~1,500, and ~1,700 cm^{-1}



True \ Predicted	Wildtype	Amikacin	Rifampicin	Isoniazid	Moxifloxacin
Wildtype	99			1	
Amikacin		99	1		
Rifampicin		1	98		1
Isoniazid	1			99	
Moxifloxacin			1	1	98

Fig. 2. BCG antibiotic resistance determination. (*A*) Average spectrum with standard deviation of all spectrograms (1,700 spectra collected at 15 s acquisition) from the five BCG strains. The spectra show significant differences in both peak position and peak intensity at bands centered at ~1,000 cm^{-1}, 1,060 cm^{-1}, and between ~1,300 to 1,700 cm^{-1}. (*B*) Two-dimensional t-SNE projection across all Raman spectra of the dataset for five BCG strains susceptible and resistant to the four different antibiotics tested grouped according to antibiotic class showing clustering. (*C*) Normalized confusion matrix generated using CNN on the few-to -single - cell spectra collected from five BCG strains, grouped by antibiotic class. Samples were evaluated by performing a stratified K-fold cross - validation of our classifier's performance across 10 splits, showing ~98% classification accuracy across all samples. (*D*) Feature selection highlighting the identification of true, physical vibrational modes. Feature selection performed to determine relative weight of spectral wavenumbers in our CNN classification. (*E*) Average classification accuracy as a function of the number of top spectral features selected showing 96% accuracy with only 500 features.

bands between wild type and four resistant strains; the majority of these vibrational modes correspond to cell wall components such as mycolic acid and proteins. Therefore, the differences between the spectra are most likely due to differences in cell-wall composition and structure, discussed later in the text. We performed t-distributed stochastic neighbor embedding (t-SNE) which shows significant clustering for each of the five strains (Fig. 2*B*). Such clustering suggests that accurate classification by resistance should be possible. To classify the antibiotic resistance profiles of the five BCG strains, we developed a deep learning model, a ResNet (25) variant (8). We perform a stratified K- fold cross- validation of our convolutional neural network's performance across 10 splits. As seen in Fig. 2*C*, we achieve an average accuracy of 98% across all BCG strains. Note that main diagonal elements represent correct classifications and off-diagonal elements represent misclassifications. High classification accuracies were also obtained with different gold- coated glass substrates (*SI Appendix*, Fig. S3) as well as with combined datasets from different experimental replicates (*SI Appendix*, Figs. S11 and S12). These results demonstrate that our approach is agnostic to the particular substrate selected or batch effects and are solely from intrinsic differences in the strains themselves. Therefore, Raman spectroscopy in conjunction with machine learning can be used to predict antibiotic resistance in mycobacteria without the need for antibiotic incubation and hence without culturing.

Select Spectral Bands Are Most Crucial for Predicting Antibiotic Resistance. We quantitatively identify the regions of the spectra that are key for the high classification accuracy. To identify these meaningful bands, we used a probing algorithm that perturbs the input data before rerunning the classification through the trained models from Fig. 2*C* (26). Using each test fold from our 10-fold cross- validation, we iterate through the wavenumbers and, at each iteration, perturb the spectrum by modulating the amplitude of the spectral intensity with a Voigt distribution centered at the probing wavenumber. After each perturbation, we recalculate the classification accuracy, compare the updated results with our baseline classification accuracy, and determine the importance for each wavenumber—the greater the decrease in accuracy due to a given perturbation, the more important the wavenumber. As shown in Fig. 2*D*, these regions are spectral bands located around 1,295, 1,437, and 1,660 cm^{-1} (27, 28). These wavenumber bands correspond to mycolic acid $-CH_2$ twist and $-CH_2$ deformation modes, and Amide I C=O stretching modes, which change when the bacteria develop resistance. Therefore, our analysis suggests that Raman spectral features can be used to differentiate resistant bacteria that exhibit distinct biochemical profiles.

Next, we reduced the wavenumber feature input to our neural network, reflecting those bands most important for high- accuracy classification. Using only the top 50 wavenumbers, we achieve ~86% accuracy. Classification accuracy increases to 96% when using 250 features, and further to 98% when using all 1,930 features (Fig. 2*E*). This result indicates potential for more rapid spectral collection and analysis as full spectroscopic analysis is not necessary to achieve high classification accuracies.

Antibiotic-resistant Mutants Can Be Accurately Classified. To test the robustness of our approach to potential patient–patient variation, we interrogated four distinct isoniazid-resistant mutants and three distinct amikacin- resistant mutants (*SI Appendix*, Table S2). As seen in Fig. 3*A*, the isoniazid-resistant mutants differ from each other by single-point mutations in the katG gene, which encodes for catalase peroxidase, an enzyme that converts isoniazid to its biologically active form (29). Isoniazid-resistant mutant 1 has a single- point mutation at position 609 from cytosine

A

Sub variant strain	Gene of interest	Mutation	Positi
Isoniazid 1	katG	C → T	609
Isoniazid 2	katG	A → G, C → T	290, 6
Isoniazid 3	katG	G → A	129
Isoniazid 4	katG	None	Non

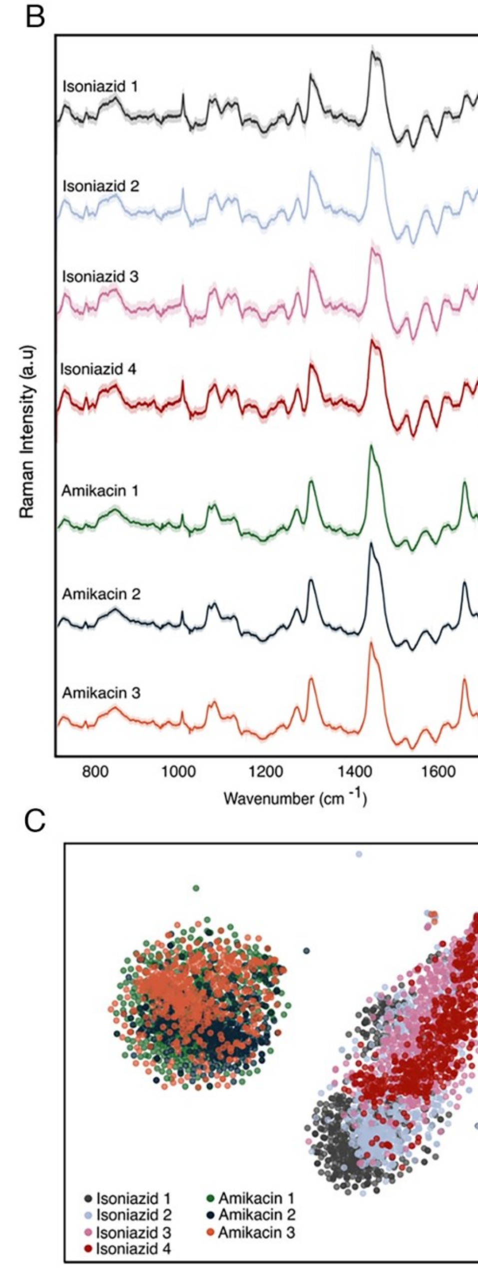


Fig. 3. Resistant mutant prediction. (*A*) Sequencing results showing single mutation differences in resistance genes of interest resulting in mutants of resistant strains against isoniazid and amikacin. (*B*) Average Raman spectra of 529 spectra per strain of four isoniazid-resistant mutants and three amikacin- resistant mutants showing nearly identical fingerprints among their respective class of resistance. (*C*) t- SNE plot showing overlapping clusters of mutants within a class of antibiotic resistance.

to thymine; isoniazid-resistant mutant 2 has two single - point mutations at positions 290 and 609 from adenine to guanine and cytosine to thymine, respectively; isoniazid-resistant mutant

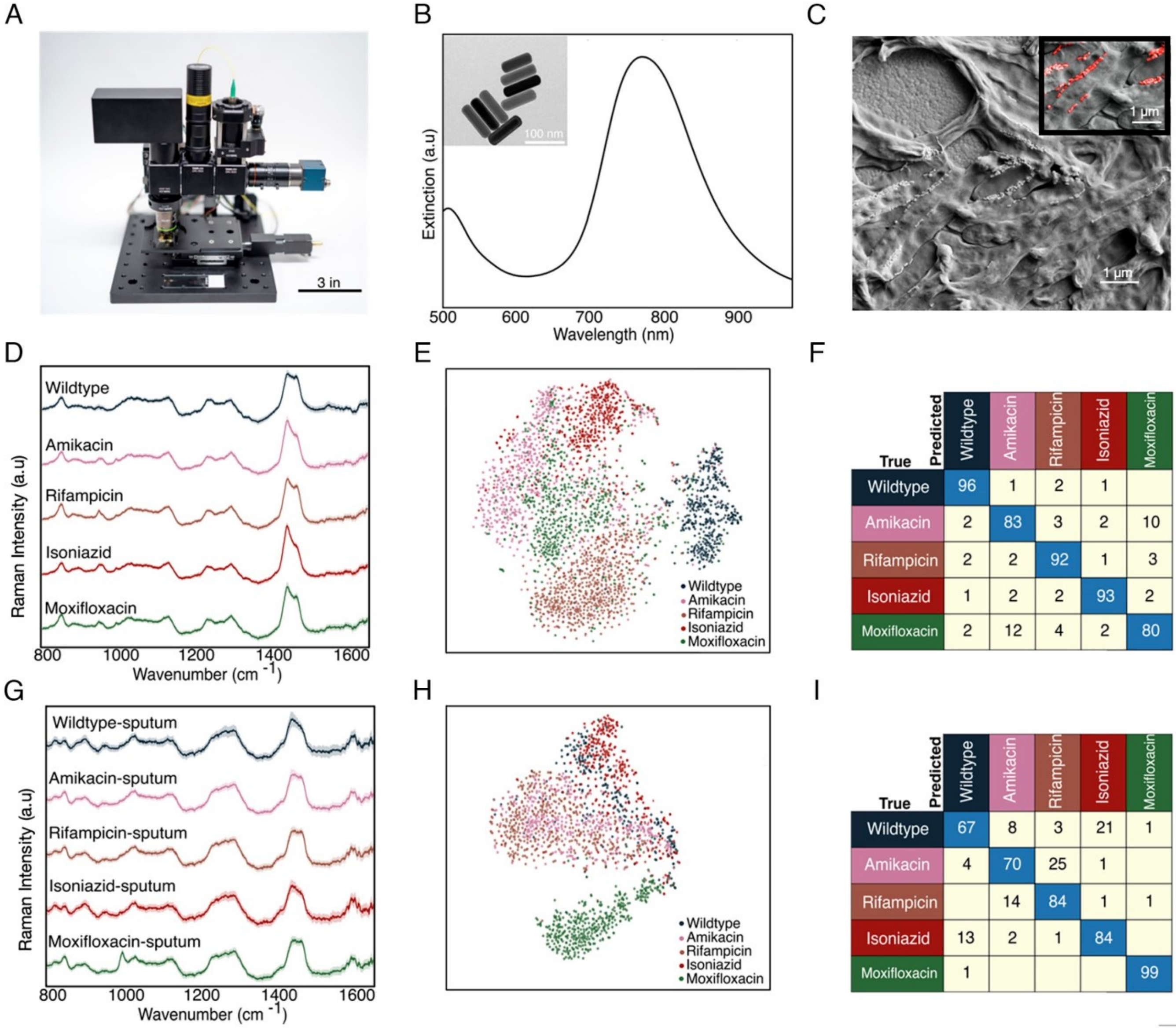


Fig. 4. Demonstration on our low-cost Raman system. (*A*) Photo of our system showcasing its small footprint and portable platform for use at the point- of- care. A standard microscope slide is included for scale. (*B*) Extinction spectra of gold nanorods showing plasmon resonance centered at ~780 nm to overlap with the setup's 785 nm excitation laser. Panel (*B*) *Inset* shows transmission electron micrograph of gold nanorods used. (*C*) Scanning electron micrograph of BCG mixed with gold nanorods, which increase the Raman scattering from single cells. Panel (*C*) *Inset* shows a false- colored scanning electron micrograph of gold nanorods colored red. (*D*) Raman spectra of the five BCG strains mixed with gold nanorods imaged on our setup. (*E*) t- SNE projection across all Raman spectra of the five BCG strains. (*F*) Average accuracy of ~89% is achieved using spectra from bacteria suspended in water and drop cast dried (replication of Fig. 2*C* on our setup). (*G*) Raman spectra of the five BCG strains spiked in sputum imaged on our setup. (*H*) t- SNE projection across all spiked sputum samples. (*I*) Average accuracy of ~79% is achieved using spectra collected from two separate spiked sputum samples.

3 has a single-point mutation at position 1,292 from guanine to adenine; and isoniazid-resistant mutant 4 has no single- point mutations identified in katG and likely harbors a non-katG mutation conferring the observed isoniazid resistance. As seen in Fig. 3*B*, Raman spectra from these distinct isoniazid- resistant mutants are seemingly identical, with tightly overlapping projections on t-SNE clustering (Fig. 3 *C*).

The three amikacin- resistant mutants reveal similar clustering on the t- SNE projection. Importantly, however, the clustering is distinct from the isoniazid cluster. The classification accuracy to predict the antibiotic susceptibility of each strain within its antibiotic class is ~99% (*SI Appendix*, Fig. S2). Clinically, this result is significant, as it is likely that antibiotic- resistant mutants from different patients may differ genetically, and we are able to identify resistance to a particular antibiotic regardless of this difference.

Enabling Low-cost and Portable Raman for Mtb Drug Testing. To demonstrate that our findings can translate to point-of - care TB drug testing in resource- limited regions, we developed a low- cost, portable, and fully automated Raman microscope (Fig. 4*A*). This Raman microscope is based on the Octopi/Squid modular microscope framework developed by our team for fluorescence (30, 31) but advances this setup for spectroscopy. The instrument utilizes a noncooled CMOS camera (Sony IMX290), a transmission diffraction grating with >95% efficiency, and a design without a pinhole or slit for low-cost spectroscopy. A fiber -coupled 785 nm VHG- stabilized laser is used as the excitation source. To boost the signal from bacteria, we synthesize gold nanoparticles (AuNPs) for surface- enhanced Raman scattering. The Dionne Group has done substantial research using gold nanorods to enhance Raman signatures from bacterial and mammalian cells for both dried and liquid samples, showing that a large range of plasmon resonances from 670 to 860 nm can be used to obtain high- quality SERS spectra for 785 nm excitation (12, 13).

As seen in Fig. 4*B*, the nanorods used in this study have a scattering resonance near the pump- laser wavelength. Compared to whole-cell Raman, these nanorods yield a ~2 order of magnitude increase in Raman scattering (10, 12, 13, 32–36) from single cells and pathogens and considerably shorten the integration times required for few-to -single - cell analysis on our microscope. The AuNPs are also relatively inexpensive to synthesize or purchase, easy to tune optically (37–39) and easy to integrate with our bacterial strains. All components combined for this microscope and spectrometer are <$5,000. Additional details on the sample preparation and imaging/spectroscopy are provided in Materials and Methods.

As a baseline measurement, we first collected ~3,000 spectra from the five BCG strains mixed with gold nanorods using 0.3 s integration times and three accumulations per sample (Fig. 4*B*). Here, the shortened integration time was chosen to maintain a similar signal- to- noise ratio of the signal as with the Raman of the scientific-grade instrument. Similar to the spectra collected on the scientific grade Raman microscope in Fig. 2*A*, the average Raman spectra of the five distinct BCG strains look similar, with major spectral peaks preserved across the strains. However, there are some noticeable differences in the ~1,000 cm^{-1}, ~1,300 cm^{-1}, and ~1,400 cm^{-1} bands between wildtype and four resistant strains (Fig. 4*D*). We performed t-SNE which shows significant clustering for each of the five strains (Fig. 4*E*). As seen in Fig. 4*F*, we achieve an average accuracy of 89% across all five BCG strains, comparable with the ~95 to 98% average classification accuracy achieved on our scientific-grade tool.

As an initial evaluation of the performance for clinical samples, we spiked BCG pathogens into sputum (*Materials and Methods*), mixed the sample with gold nanorods, and collected a total of ~5,300 spectra from the BCG strains. Using only 0.3 s integration times and three accumulations per sample, we obtain an average classification accuracy of ~79% for the five strains across two separate sputum samples (Fig. 4*I* and *SI Appendix*, Fig. S7). This accuracy is reasonable with that observed for bacteria alone on scientific-grade Raman instruments (Fig. 2 *C*) while providing rapid results at a lower cost and smaller footprint.

We note that the spectra differ from those without using AuNPs, which is as expected as now the Raman signatures primarily originate from the surface-bound molecules on the bacteria, due to the electric field localization of the AuNPs. This intraclass spectral variation is evident in the milder, less intense clustering of the five strains in the t- SNE in Fig. 4 *E* and *H* compared to the whole-cell Raman collected in Fig. 2*B*. However, our convolution neural network can still classify strains, as demonstrated by the high accuracies we are able to obtain. Future work spanning more patient data and optimized spectral data processing could enable clinical BCG strain classification.

Discussion

In summary, we demonstrate a rapid and accurate TB AST methodology that is antibiotic incubation-free and culture - free. Unlike current culture-based and nucleic acid amplification - based testing platforms, our Raman approach does not require lengthy culturing, expensive reagents, or thermal cycling equipment, and is robust to single nucleotide point mutations in resistance genes of interest. Our approach can accurately classify resistance across a variety of antibiotics and across a variety of genetic mutations. In addition, our machine learning models not only provide accurate AST classification, but also interpretability to the Raman spectra. Indeed, as highlighted by our feature- recognition algorithm, resistance behavior is primarily reflected in certain vibrational modes, particularly those of mycolic acid, that underly the molecular- level antibiotic response of the bacteria. We developed a low-cost portable Raman microscope, capable of high BCG classification accuracy on dried samples and in dried sputum. To extend these results beyond that of the BCG laboratory strain, on-going work is aimed at testing our approach in the field on Mtb-positive patients. Specifically, we are planning to fold Octopi-Raman deployment into the growing Octopi infrastructure, which currently includes deployment all across the globe, including India, Nigeria, Senegal, Liberia, Kenya, and the United Kingdom. Such clinical work will elucidate the intrinsic differences between strain types, well beyond sputum-to - sputum variations. If successful, this clinical work will provide a needed approach to enable the TB surveillance and diagnostic milestones set by the WHO. Finally, our study does have limitations: Our work is a proof of concept that probes the phenotypic differences of singly antibiotic- resistant laboratory BCG strains using Raman spectroscopy. However, further work should be done using this methodology to probe multidrug- resistant BCG and Mtb. With more work in this direction, our methodology could be applied to curbing the growing global crisis of multidrug- resistant TB.

Materials and Methods

Bacterial Strains. To raise antibiotic- resistant mutants, BCG was plated on 7H10 media with increasing concentrations of each antibiotic of interest. Plates were incubated at 37 °C for 21 d. Individual colonies were selected and grown in 7H9 media containing OADC, 0.2% glycerol, and 0.05% Tween- 80; antibiotic resistance was confirmed by MIC testing.

MIC Determination. Each strain was grown to mid- log phase in 7H9 media containing OADC, 0.2% glycerol, and 0.05% Tween- 80. Cultures were then diluted back to an OD600 of 0.01 in media containing twofold dilutions of each antibiotic of interest. Cultures were grown at 37 °C in 96-well plates; bacterial cells were mixed at 7 and 14 d of incubation, and growth was determined by OD600 measurements on a Tecan Spark plate reader.

Bacterial Preparation for Spectroscopy. To obtain the antibiotic- resistant BCG strains, wild-type BCG [ATCC 35737 TMC 1019 (BCG

Japanese)] was grown to mid - log phase in 7H9 liquid media supplemented with OADC, glycerol, and tween- 80, then pelleted and plated on 7H10 solid media containing serial dilutions of each antibiotic of interest from 10- fold above the MIC to 10- fold below the MIC. Plates were incubated at 37 °C for 4 wk; individual colonies were selected for culture from the serial dilutions that yielded between 1 and 25 colonies. One wild- type colony was selected from the wildtype plate, one rifampicin-resistant colony was selected from the rifampicin plate, four amikacin-resistant colonies were selected from the amikacin plate, five isoniazid-resistant colonies were selected from the isoniazid plate, and five moxifloxacin-resistant colonies were selected from the moxifloxacin plate. Each colony had its antibiotic resistance profile confirmed using MIC and each colony was confirmed to be unique using whole-genome sequencing and MIC and hence was used as substrains (*SI Appendix*, Table S2).

To determine MIC, wild-type BCG and each resistant strain were grown to mid - log phase, then diluted back to an OD600 of 0.01 and incubated in twofold serial dilutions of each antibiotic in a 96-well plate. On days 7 and 14 postinoculation, each well was gently mixed and OD600 was read on a TECAN Spark device. Only confirmed resistant mutants as well as the fully susceptible (wildtype) strains were used for Raman testing. One mL of frozen stock of the wild type and the confirmed resistant mutants were seeded in 10 mL 7H9 liquid media supplemented with OADC, glycerol, and tween- 80 and incubated under 37 °C shaking at 115 rpm until mid-log phase (~5 to 6 d) using the Thermo Scientific MaxQ 4450 incubator. Cultures were then chemically inactivated overnight in a 1:1 solution using 10% formalin. Then, 1.5 mL culture [~4 × 10 (8) cells/mL] was washed with water three times at 6,000 rpm for 4 min using a mySPINTM 6 Mini Centrifuge. The BCG strains were prepared for Raman interrogation by resuspending the final washed pellet in 100 μL of sterile water and drying 3 μL of the suspension on a gold- coated glass substrate. Droplets were dried for 20 min, and Raman was collected immediately after drying.

For the sputum- spiked BCG Raman interrogation, 25 μL of 3× washed bacteria was spun down and resuspended in 50 μL of sputum. After sitting for 1 min, the spiked sputum was then spun down, washed once with DI water, and then resuspended in 25 μL of DI water. Then, 5 μL of cell solution was mixed with nanorods in a 1:1 ratio, and 3 μL of the suspension was dried on a gold-coated glass substrate.

Raman Spectral Data Collection and Processing. The BCG strains were illuminated with a 633 nm laser powered at 13.17 mW. The Raman spectra were collected from 742 cm^{-1} to 1,802 cm^{-1} using a 100×, 0.9 NA objective with a spot size of ~3.5 μm using the scale provided in the Horiba software, a 600 l/mm grating, and 15-s acquisition time per spectra, ensuring that spectra are collected from single to at most a few bacterial cells at a time. Python (Jupyter Notebook) was used to process spectral data. Individual spectra were smoothed using wavelet denoising, specifically using the denoise wavelet function from the scikitimage Python library. Background was then subtracted using a polynomial fit with degree 5, and spectra were individually normalized to have zero mean and unit variance across all wavenumbers.

Fabrication of Gold- Coated Glass Substrates. The gold- coated glass substrates used in this work were prepared by evaporating a 5 nm adhesion layer of titanium, followed by 200 nm of gold at a rate of 1 A/s using the Kurt J. Lesker E- Beam Evaporator in the Stanford Nano Shared Facilities. Gold was deposited using a low deposition rate of 1 Å/s to ensure high uniformity and quality and minimal roughness throughout the thin film substrate. The gold film helps minimize the background fluorescence signal from the underlying glass substrate.

Gold Nanorod Preparation for SERS Measurement. Hexadecyl(trimethyl) ammonium bromide (CTAB) and sodium oleate (NAOL)- coated gold nanorods were synthesized following previously described protocol (37, 40). The nanorods were cleaned by centrifuging 1 mL aliquots once at (9,000 rpm, 20 min) and were concentrated down to 50 μL volumes. Nanorods were mixed with bacteria in 1:1 volumetric ratio to enable uniform coating as described in our earlier work (12). Absorption spectra were recorded using a Cary 5000 UV- vis- NIR spectrometer. Scanning electron microscopy images were taken using FEI Magellan 400 XHR scanning electron microscope and the Zeiss Sigma SEM. Transmission electron microscopy images were taken using FEI Tecnai G2 F20 X-TWIN transmission electron microscope (TEM). Gold nanorods with a plasmon resonance centered at 780 nm with a height of ~99 nm and a width of ~28 nm were used.

CNN Architecture and Classification. Our CNN model is adapted from a ResNet (25) architecture, consisting of an initial convolutional layer with 64 filters, seven residual layers, and one fully connected layer. Each residual layer consists of two residual blocks and each convolutional layer has 32 hidden nodes. We used a minibatch size of 32. To assess the classification performance, we used 10-fold cross -validation, which splits the dataset into 10 equal stratified sets (Scikit- learn StratifiedKFold). For each fold i, we leave set i as the test set and train model i using all remaining data. Model i is then used to classify data in set i (test set). We repeat this process for all 10- fold, which produces unbiased classifications for all the data, and these are compared against the ground-truth to generate confusion matrices and the overall accuracy. All t-SNE projections were plotted using Scikit- learn manifold t-SNE with a perplexity =15.

To determine the most important wavenumbers or spectral bands affecting classification performance, we used a probing algorithm that perturbs the input data before rerunning the classification through the trained models. Using the test fold from our 10- fold cross- validation, we iterate through the wavenumbers and, at each iteration, perturb the spectrum by modulating the amplitude of the spectral intensity with a Voigt distribution centered at the probing wavenumber. After each perturbation, we recalculate the classification accuracy, compare the updated results with our baseline classification accuracy, and determine the importance for each wavenumber—the greater the decrease in accuracy due to a given perturbation, the more important the wavenumber. Each wavenumber of each spectrum in the test fold is perturbed five times by randomly sampling spectral intensity values at that wavenumber in order to create random, but spectrally sensible instances. Finally, all results are averaged to determine our final feature importance.

Data Collection on Octopi- Raman. Raman spectra were collected from 350 cm^{-1} to 1,606 cm^{-1} using a 20×, 0.75 NA objective with spot size of 5.5 μm (Gaussian diameter; FWHM of 3.2 μm), laser power of 2.53 mW, and 0.3 second acquisition time per spectrum. The laser spot size was determined using a stage micrometer as a standard.

Sputum Samples. To obtain processed sputum for Raman interrogation, sputum samples were decontaminated and concentrated per standard laboratory procedure (41) as follows: the sputum samples were treated with N-acetyl -l - cysteine–sodium hydroxide (NALC- NaOH) to break up the mucus and treated with NaOH to kill the oral flora. The pH was then neutralized with phosphate buffer and samples were concentrated with centrifugation. Only sputa with a negative acid- fast bacilli microscopy result (mycobacteria-negative) and from patients at low risk for TB were used for BCG spiking. Patient samples were deidentified prior to use in our study.

Data, Materials, and Software Availability. Raman spectra of BCG strains from the datasets used in this paper are deposited in Github (42).

ACKNOWLEDGMENTS. We thank Chris Cundy, Ahmed Shuabi, Dr. Fareeha Safir, Dr. Jack Hu, Dr. Halleh Balch, and Liam Herndon for many insightful discussions. We thank Dr. Fareeha Safir for assistance with collecting SEM images of Bacillus Calmette–Guérin (BCG), Liam Herndon for assistance with gold nanorod synthesis, and Ariel Stiber for assistance collecting TEM images of gold nanorods. We thank the Chan Zuckerberg Biohub San Francisco for funding, as well as the Stanford BMGF- funded Institute for Immunity, Transplantation, and Infectious Disease for seed funding. We would also like to thank Triet Doan for assistance with BCG DNA isolation and Vida Shokoohi and Dr. John Coller at the Stanford Functional Genomics Facility for their assistance with BCG sequencing. Geneious Prime 2024.0.2 was used to analyze the BCG sequencing data. Some elements of Fig. 1*A* were created with BioRender.com. We acknowledge the support of the following funding sources: Stanford Catalyst for Collaborative Solutions (funding ID 132114), Stanford BMGF-funded Institute for Immunity, Transplantation, and Infectious Disease (grant number OPP 1113682), The Chan Zuckerberg Biohub San Francisco Investigator Program, The Gates Foundation (grant number OPP 1113682), The National Science Foundation (grant number 1905209), The NIH New Innovator Award (1DP2AI152072-01) , the DARE Fellowship (L.F.T.), the Stanford Graduate Fellowship (B.O.), the Stanford EDGE Fellowship (B.O.), the Stanford Bio- X SIGF Paul Berg Interdisciplinary Biomedical Graduate Fellowship (H.L.) and Stanford Bio- X Bowes Graduate Student Fellowship (B.O.). Part of this work was performed at the Stanford University Nano Shared Facilities and Stanford University Nanofabrication Facilities, which are supported by the National Science Foundation and National Nanotechnology Coordinated Infrastructure under awards ECCS-2026822 and ECCS-1542152 as well as the Stanford University Cell Sciences Imaging Core Facility.

Author affiliations: [a]Department of Materials Science and Engineering, Stanford
University, Stanford, CA 94305; [b]Department of Bioengineering, Stanford University
School of Medicine and School of Engineering, Stanford, CA 94305; [c]Department of Mechanical Engineering, Massachusetts Institute of Technology, Cambridge, MA 02142; [d]The Ragon Institute of Mass General, Massachusetts Institute of Technology, and Harvard, Cambridge, MA 02139; [e]Jameel Clinic for AI & Healthcare, Massachusetts Institute of Technology, Cambridge, MA 02139; [f]Department of Electrical Engineering, Stanford University, Stanford, CA 94305; [g]Pumpkinseed Technologies, Inc., Palo Alto, CA 94306; [h]Department of Pathology, Stanford University School of Medicine, Stanford, CA 94305; [i]Division of Infectious Diseases, Massachusetts General Hospital, Boston, MA 02114; [j]Department of Medicine, Harvard Medical School, Boston, MA 02115; [k]Department of Engineering Mathematics and Physics, Cairo University, Faculty of Engineering, Giza 12613, Egypt; and [l]Department of Radiology, Molecular Imaging Program at Stanford, Stanford University School of Medicine, Stanford, CA 94035

Author contributions: B.O., L.F.T., H.L., A.K.B., A.A.E.S., M.P., and J.A.D. designed research; B.O., L.F.T., and H.L. performed research; N.V., N.B., A.K.B., and M.P. contributed new reagents/analytic tools; B.O., L.F.T., H.L., and N.V. analyzed data; and B.O., L.F.T., H.L., N.V., A.K.B., A.A.E.S., and J.A.D. wrote the paper.

1. "Global tuberculosis report 2022" (WHO Tech. Rep. 2023).
2. "WHO consolidated guidelines on tuberculosis: module 3: diagnosis: rapid diagnostics for tuberculosis detection, 2021 update" (WHO Tech. Rep. 2021).
3. S. D. Lawn *et al.*, Advances in tuberculosis diagnostics: The Xpert MTB/RIF assay and future prospects for a point- of- care test. *Lancet Infect. Dis.* **13**, 349–361 (2013).
4. A. Barczak *et al.*, RNA signatures allow rapid identification of pathogens and antibiotic susceptibilities. *Proc. Natl. Acad. Sci. U.S.A.* **109**, 6217–6222 (2012).
5. World Health Organisation, High priority target product profiles for new tuberculosis diagnostics: Report of a consensus meeting (World Health Organisation, Geneva, 2014).
6. M. Varona, J. L. Anderson, Advances in mutation detection using loop-mediated isothermal amplification. *ACS Omega.* **6**, 3463–3469 (2021).
7. D. W. Shipp, F. Sinjab, I. Notingher, Raman spectroscopy: Techniques and applications in the life sciences. *Adv. Opt. Photonics.* **9**, 315 (2017).
8. C.- S. Ho *et al.*, Rapid identification of pathogenic bacteria using Raman spectroscopy and deep learning. *Nat. Commun.* **10**, 4927 (2019).
9. S. Stöckel *et al.*, Raman spectroscopic identification of Mycobacterium tuberculosis. *J. Biophotonics.* **10**, 727–734 (2017).
10. A. A. Moawad *et al.*, A machine learning-based raman spectroscopic assay for the identification of Burkholderia mallei and related species. *Molecules.* **24**, 4516 (2019).
11. M. Dahms *et al.*, Raman spectroscopic differentiation of *Streptococcus pneumoniae* from other Streptococci using laboratory strains and clinical isolates. *Front. Cell Infect. Microbiol.* **12**, 930011 (2022).
12. L. F. Tadesse *et al.*, Plasmonic and electrostatic interactions enable uniformly enhanced liquid bacterial surface- enhanced raman scattering (SERS). *Nano Lett.* **20**, 7655–7661 (2020).
13. F. Safir *et al.*, Combining acoustic bioprinting with AI-assisted Raman spectroscopy for high - throughput identification of bacteria in blood. *Nano Lett.* **23**, 2065–2073 (2023).
14. K. Dragounova *et al.*, Identification of bacteria in mixed infection from urinary tract of patient's samples using Raman analysis of dried droplets. *Analyst* **148**, 3806–3816 (2023).
15. S. Khan *et al.*, Analysis of hepatitis B virus infection in blood sera using Raman spectroscopy and machine learning. *Photodiagnosis Photodyn. Ther.* **23**, 89–93 (2018).
16. E. Kastanos, A. Kyriakides, K. Hadjigeorgiou, C. Pitris, A novel method for bacterial UTI diagnosis using Raman spectroscopy. *Int. J. Spectrosc.* **2012** (2012), https://www.hindawi.com/ archive/2012/195317/abs/.
17. A. Germond *et al.*, Raman spectral signature reflects transcriptomic features of antibiotic resistance in Escherichia coli. *Commun. Biol.* **1**, 85 (2018).
18. J. Ye *et al.*, Accurate virus identification with interpretable Raman signatures by machine learning. *Proc. Natl. Acad. Sci. U.S.A.* **119**, e2118836119 (2022).
19. J. Plou *et al.*, Machine learning-assisted high -throughput SERS classification of cell secretomes. *Small* **19**, e2207658 (2023).
20. A. N. Rousseau *et al.*, Fast antibiotic susceptibility testing via raman microspectrometry on single bacteria: An MRSA case study. *ACS Omega.* **6**, 16273–16279 (2021).
21. W. R. Premasiri *et al.*, Rapid urinary tract infection diagnostics by surface-enhanced Raman spectroscopy (SERS): Identification and antibiotic susceptibilities. *Anal. Bioanal. Chem.* **409**, 3043–3054 (2017).
22. B. L. Thomsen *et al.*, Accurate and fast identification of minimally prepared bacteria phenotypes using Raman spectroscopy assisted by machine learning. *Sci. Rep.* **12**, 16436 (2022).
23. B. Kaewseekhao *et al.*, Diagnosis of active tuberculosis and latent tuberculosis infection based on Raman spectroscopy and surface- enhanced Raman spectroscopy. *Tuberculosis* **121**, 101916 (2022).
24. L. Wang *et al.*, Machine learning analysis of SERS fingerprinting for the rapid determination of
Mycobacterium tuberculosis infection and drug resistance. *Comput. Struct. Biotechnol. J.* **22** (2022).
25. K. He, X. Zhang, S. Ren, J. Sun, "Deep residual learning for image recognition" in *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2016), pp. 770–778.
26. F. Safir *et al.*, Combining acoustic bioprinting with AI-assisted Raman spectroscopy for high - throughput identification of bacteria in blood. *Nano Lett.* **23**, 2065–2073 (2023).
27. A. Zyubin *et al.*, Raman spectroscopy reveals M. tuberculosis strains with different antibiotic susceptibility. *Laser Phys. Lett.* **16**, 085602 (2019).
28. T.- Y. Liu *et al.*, Differentiation of bacteria cell wall using Raman scattering enhanced by nanoparticle array. *J. Nanosci. Nanotechnol.* **12**, 5004–5008 (2012).
29. V. R. Bollela *et al.*, Detection of katG and inhA mutations to guide isoniazid and ethionamide use for drug- resistant tuberculosis. *Int. J. Tuberc. Lung Dis.* **20**, 1099–1104 (2016).

30. H. Li, H. Soto- Montoya, M. Voisin, L. F. Valenzuela, M. Prakash, Octopi: Open configurable high- throughput imaging platform for infectious disease diagnosis in the field. bioRxiv [Preprint] (2019). https://doi.org/10.1101/684423 (Accessed 1 August 2022).
31. H. Li *et al.*, Squid: Simplifying quantitative imaging platform development and deployment. bioRxiv [Preprint] (2020). https://doi.org/10.1101/2020.12.28.424613 (Accessed 26 November 2022).
32. J.- Y. Lim *et al.*, Identification of newly emerging influenza viruses by surface-enhanced raman spectroscopy. *Anal. Chem.* **87**, 11652–11659 (2015).
33. R. Adato *et al.*, Ultra- sensitive vibrational spectroscopy of protein monolayers with plasmonic nanoantenna arrays. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 19227–19232 (2009).
34. M. Bhamidipati, H.- Y. Cho, K.- B. Lee, L. Fabris, SERS- based quantification of biomarker expression at the single cell level enabled by gold nanostars and truncated aptamers. *Bioconjug. Chem.* **29**, 2970–2981 (2018).
35. Y.-C. Ou *et al.*, Diagnosis of immunomarkers in vivo via multiplexed surface enhanced Raman spectroscopy with gold nanostars. *Nanoscale.* **10**, 13092–13105 (2018).
36. R. Zhang *et al.*, Chemical mapping of a single molecule by plasmon-enhanced Raman scattering. *Nature.* **498**, 82–86 (2013).
37. N. R. Jana, L. Gearheart, C. J. Murphy, Wet chemical synthesis of high aspect ratio cylindrical gold nanorods. *J. Phys. Chem. B.* **105**, 4065–4067 (2001).
38. P. K. Jain, K. S. Lee, I. H. El- Sayed, M. A. El- Sayed, Calculated absorption and scattering properties of gold nanoparticles of different size, shape, and composition: Applications in biological imaging and biomedicine. *J. Phys. Chem. B.* **110**, 7238–7248 (2006).
39. K.- Q. Lin *et al.*, Size effect on SERS of gold nanorods demonstrated via single nanoparticle spectroscopy. *J. Phys. Chem. C.* **120**, 20806–20813 (2016).
40. X. Ye, C. Zheng, J. Chen, Y. Gao, C. B. Murray, Using binary surfactant mixtures to simultaneously improve the dimensional tunability and monodispersity in the seeded growth of gold nanorods. *Nano Lett.* **13**, 765–771 (2013).
41. K. C. Carroll *et al.*, *Manual of Clinical Microbiology, 2 Volume Set* (Wiley, 2019).
42. B. Ogunlade, L. Tadesse, H. Li, Rapid- antibiotic- incubation- free- determination-of- *tuberculosis*- drug- resistance- using- ML- and- Raman. Github. https://github.com/bogun98/Rapid- antibiotic- incubation- free- determination-of- tuberculosis- drug- resistance- using- ML- and- Raman. Deposited 29 April 2024.