Profiles of Volatile Biomarkers Detect Tuberculosis from Skin

Hossam Haick (hhossam@technion.ac.il)
Technion – Israel Institute of Technology
Rotem Vishinkin
Technion - Israel Institute of Technology https://orcid.org/0000-0001-7799-5550
Rami Busool
Technion – Israel Institute of Technology https://orcid.org/0000-0003-0598-0127
Elias Mansour
Technion – Israel Institute of Technology
Falk Fish
Technion – Israel Institute of Technology
Aliasgar Esmail
University of Cape Town
Parveen Kumar
All India Institute of Medical Sciences
Alaa Garaa
Technion – Israel Institute of Technology
John Cancilla
Scintillon Institute
Jose Torrecilla
Complutense University of Madrid
Girts Skenders
University of Latvia
Marcis Leja
University of Latvia
Keertan Dheda
University of Cape Town https://orcid.org/0000-0001-7709-5341
Sarman Singh
All India Institute of Medical Sciences, Bhopal https://orcid.org/0000-0002-0749-9647

Article

Keywords: Bio-diagnostic Tests, Physiological Pathway, Volatile Organic Compounds, Skin Headspace, Active Pulmonary TB patients

Posted Date: December 10th, 2020

DOI: https://doi.org/10.21203/rs.3.rs-119729/v1

Abstract

Tuberculosis (TB) is an infectious disease that threatens >10 million people annually. Despite advances in TB diagnostics, millions of patients continue to receive insufficient diagnosis, as TB symptoms are not specific. Many existing bio-diagnostic tests are slow, have low sensitivity and/or specificity, and can be too expensive or complex for resource-limited settings. Early diagnosis of TB contributes to morbidity mitigation and prevents the disease spread. Here, we report on a new physiological pathway for TB diagnostics via TB-specific profile of volatile organic compounds (VOCs) that are detected and quantified from air trapped above the skin ("skin headspace"). Beyond the exploration aspect that relies on clinical samples from 636 individuals in two clinical sites, we demonstrate the utility of these VOC profiles in a point-of-care diagnosis, by means of specifically-designed intelligent nanoarray, both in lab setting and online, in-situ wearable devices. The results show excellent discrimination between active pulmonary TB patients and controls with 89.4% accuracy. This fulfills the triage TB test requirements according to the World Health Organization (WHO).

Introduction

Tuberculosis (TB) is a major health problem in the world.[1,2] Approximately 95% of TB cases occur in developing countries, including locations where people live on less than 1 USD per day. About one-third of the world population has latent TB with a lifetime risk of 5 to 10% of developing active TB.[3] HIV co-infection, smoking and malnutrition greatly increase this risk and speed up the TB epidemic.[1,2] TB is particularly difficult to diagnose in children and in co-infected populations with HIV.[4] Currently, around 3 million active TB cases are missed by the health systems worldwide.[5] Despite advances in TB diagnostics, millions of patients continue to receive incomplete or delayed diagnosis, as the physical signs and symptoms of TB are non-specific.[6]

Many existing bio-diagnostic tests are available. Nevertheless, they are mostly slow, have low sensitivity and/or specificity, and, at times, are too expensive or complex for resource-limited settings. For example, a sputum smear (2.6 to 10.50 USD/examination, depending on the country) is too insensitive, and mycobacterial culture takes 4-8 weeks and at least 3 visits by the patient to finalize the diagnosis and begin treatment.[6] This process is time-consuming, labor intensive, requires highly trained technicians and the method is based on challenging specimen collection and processing, both of which can greatly affect the sensitivity. Despite the high specificity, direct smear microscopy is relatively insensitive (20-80% sensitivity), since at least 5,000 bacilli per milliliter of sputum are required for a positive result.[7] The sensitivity is further reduced in patients with extra-pulmonary TB, HIV-compromised patients, and those with disease due to non-TB mycobacteria.[8,9] The limitations of using microscopy in low-resource settings include poor quality reagents, unmaintained microscopes, and poor level of staff training. The only technology that can diagnose within 2 hours, GeneXpert MTB/RIF, has a relatively limited sensitivity performance (88% as initial test and 67% as add-on test for TB detection following a negative smear-microscopy result).[10] The equipment and consumables required are costly, as is the initial capital cost for the GeneXpert unit and additional costs for delivery, installation and servicing. The GeneXpert test is focused on sputum as the sample, and cannot differentiate between live and dead bacteria.[11] Disadvantages of other methods include sputum as the sampling method, a requirement of a highly trained staff, high biosafety and maintenance levels for performance, along with high cost. Therefore, new, and accurate TB diagnostic methods that can be produced and distributed at affordable prices for people living on <1 USD/day are critically needed[12]. Early diagnosis and treatment initiation mitigate morbidity and disease spearing. From economic point of view, TB

cause about 12 billion dollars to vanish from the global economy when considering the cost of TB patients' loss of productivity and deaths cases[13].

We report in this article on the exploration and application of TB-specific volatile organic compounds (VOCs) that can be detected from air trapped above the skin (the “skin headspace”). Deviation of these VOCs from the healthy VOC pattern in terms of their concentration range may indicate either TB infection or TB high infection risk. For translating these findings into a point-of-care reality, we present and discuss a new biomedical apparatus containing a flexible and wearable polymeric pouch for the collection and storage of skin VOCs and their analysis by nanotechnology-based sensors array in conjunction with machine learning. The clinical offline study in two countries shows that this approach provides a fast, precise detection and classification of TB profiles from the patches collected from the skin’s headspace. Furthermore, as an additional step for realization towards a point of care diagnosis tool, an exploratory pilot study was conducted by applying a wearable electronic device directly on the skin.

Results

Study design and skin sampling. Our study had four off-line stages for examining the hypothesis and the science behind it and one demonstration stage with online and in-situ wearable device**.** The offline stage took place through the inclusion of 636 subjects aged 22-60 years (*see* Fig. 1): Skin sampling, Gas Chromatography Mass Spectrometry (GC-MS) analysis, and analysis by nanomaterial-based sensors in conjugation with pattern recognition methods. To create a robust tool for TB detection, samples, and analysis in two clinical were established in Cape Town in South Africa (N=320) and New Delhi in India (N=316). The study population included confirmed pulmonary active TB cases, healthy volunteers, and confirmed non-TB cases. Demographic and clinical data of the population is summarized in Tables S1 and S2 of the SI. 13 potential confounding factors, including HIV status and smoking habits, were monitored, and assessed. Headspace samples from each participant were collected from the anterior arm area (inner arm) and chest area using both: i) two offline porous polymeric pouch containing poly(2,6-diphenylphenylene oxide) polymer; and ii) two polydimethylsiloxane (PDMS) sheets[14] covered by an adhesive medical tape. As a reference, poly(2,6-diphenylphenylene oxide) pouches and PDMS sheets for room sampling were also included in the analysis to evaluate the impact of the exogenous (bio)chemical factors. Comparative analysis has shown that sampling by poly(2,6-diphenylphenylene oxide)-based sampling at the anterior arm area give best and most stable results, making it the focus of our presentation of the current article. For more details about the various skin sampling, please refer to SI, Section 2. Fig. 1a presents schematically the sampling and analysis with gas chromatography linked with mass spectrometry (GC-MS) as well as with offline sensors array in conjugation with pattern recognition patterns.

Gas Chromatography Mass Spectrometry (GC-MS) analysis. In the second stage (Fig. 1b), we identified the skin VOCs that potentially can serve as TB biomarkers and, furthermore, preformed qualitative and quantitative analyses using GC-MS. GC-MS analysis revealed 27 and 35 peaks with at least 80% non-zero values in all subjects’ skin samples in South Africa and India’s clinical sites, respectively, after discarding column bleeding compounds, such as siloxane-related ones. All the peaks presented a right-skewed distribution. The statistical evaluation was based on adjusted p-value for multiple peaks evaluation using a non-parametric Kruskal-Wallis test and a Steel method in comparison to the TB group as a post hoc testing. In Cape Town, the analysis included: (i) 89 confirmed pulmonary active TB patients; (ii) 90 non-TB patients with healthy controls; and (iii) 262 room samples. 4 VOCs

were found to be significantly different in comparison to the pulmonary active TB group and include toluene (retention time (R.T.) 8.4 min), acetic acid (R.T. 3.34 min), 2-ethyl-1-hexanol (R.T. 13.7 min) and tentatively recognized hexyl butyrate (R.T. 18.8 min). In New Delhi, the analysis was based on: (i) 89 confirmed pulmonary active TB patients; (ii) 193 non-TB patients with healthy controls; and (iii) 193 room samples. 3 VOCs were found to be significantly different in comparison to the pulmonary active TB, among them was toluene, which was also at higher levels among confirmed active TB patients as in Cape Town. Additional statistically significant VOCs included both tentatively recognized, ethyl-cyclopropane (R.T. 2.8 min) and octanoic acid (R.T. 15.23 min) compounds.

Fig. 2 and Table 1 present the information regarding each VOC in different test groups from both clinical sites. GC-MS analysis of samples from both clinical sites revealed TB-associated skin VOC profiles that differed from those of the control profiles and room samples. In general, during disease formation, cells undergo structural and metabolic changes that change VOCs patterns.[15] As a result, some of these VOCs appear in distinctive compositions of the mixture, depending on whether a cell is healthy or infected.[16–18] Higher levels of **acetic acid** among confirmed active TB patients may be evidence of the response of the immune system during infection.[19] **2-ethyl-1-hexanol** was previously reported as a TB-related VOC in the exhaled breath of patients, indicating its relevance to TB pathogenesis, strengthening the higher abundances among paitents.[20] **Hexyl butyrate** abundance was significantly higher at room samples in comparison to the skin samples and obtained the lowest abundance among non-TB group. The correlation to the TB disease is unclear; though it known to have both endogenous[21] and exogenous sources.[22,23] This compound is related to lipid metabolism pathway and its derivatives were found in the exhaled breath of healthy subjects.[21] **Toluene** appeared in significantly higher levels among confirmed pulmonary active TB patients in both clinical sites, yielding similar ratios with non-TB abundance. The increased levels suggest a role for toluene both in bacterium metabolism and the immune system during the infection.[24–28] Although some studies report on toluene emission during poly(2,6-diphenylphenylene oxide) (TENAX-TA®) degradation and storage, the conclusions are inconclusive.[29–35] Furthermore, GC-MS analysis of conditioned and unused poly(2,6-diphenylphenylene oxide) pouches after 8 months at 4°C refrigerator storage, in our laboratory, didn't contain toluene residues. These findings strengthen toluene's accountably a potential TB marker. Increased levels of **ethyl-cyclopropane** from TB subjects emphasizes the critical role of cyclopropane in the progress of infection[36] and immune activation[37]. **Octanoic acid** was found with the highest abundance in room samples and with the lowest abundance among confirmed pulmonary active TB patients. Similar to the Hexyl butyrate, octanoic acid has both endo-[21,38,39] and exogenous sources.[40] Lower abundance among confirmed pulmonary active TB patients may be correlated to damaged body's energy regulation during the illness in comparison to control subjects.[38,39] It is important to accentuate that the proposed metabolic explanations are hypotheses based on literature as serve as a starting point for further investigation. For more details on the potential biochemical sources of the determined VOCs, see SI, Section 3. Validation and quantification of the compounds were carried out by calibration curves. The identification and quantification of ethyl-cyclopropane, hexyl butyrate and octanoic acid were not achieved due to permission system limitations.

Table 1 Summary of VOCs' properties, including simulated synthetic samples for validation and quantification of each VOC.

	South African site				Indian site		
	Acetic acid	2-ethyl-1-hexanol	Hexyl butyrate	Toluene		Ethyl-cyclopropane	**Octanoic Acid**
Formula	$C_2H_4O_2$	$C_8H_{18}O$	$C_{10}H_{20}O_2$	C_7H_8		C_5H_{10}	$CH_3(CH_2)_6CO_2H$
CAS no.	64-19-7	104-76-7	2639-63-6	108-88-3		1191-96-4	124-07-2
R.T. (min)	3.34	13.70	18.80	8.40		2.82	14.96
m/z (mass to charge)	43	57	43	91		42	60
Laboratory simulations: Mean, ppb (s.e)							
Confirmed pulmonary active TB patients	936.47 (81.02)	48.18 (4.09)	-	332.68 (30.49)	539.02 (62.96)	-	-
Non-TB patients and healthy controls	722.63 (73.55)	35.56 (3.33)	-	220.86 (25.60)	394.30 (29.12)	-	-
Room samples	432.32 (34.39)	16.87 (1.90)	-	245.92 (19.19)	353.20 (30.31)	-	-
Lowest tested concentration (ppb)	700	6	-	60		-	-
p-value for subgroup comparisons[c]							
Kruskal-Wallis Test	<0.0001[a]	<0.0001[a]	<0.0001[a]	0.0015[a]	0.0003[b]	<0.0001[b]	<0.0001[b]
Confirmed pulmonary active TB patients vs. Non-TB patients and healthy controls	0.0295	0.0313	0.0078	0.0022	0.0077	0.0093	0.0060 (0.0018)[d]
Confirmed pulmonary active TB patients vs. Room	<0.0001	<0.0001	0.0341	0.0048	<0.0001	<0.0001	<0.0001

[a] α = 0.00185. [b] α = 0.00143. [c] Steel method in comparison to the TB group as a post hoc testing. [d] After elimination of two extreme points.

Nanomaterial-based sensors in conjugation with pattern recognition analysis. In the third stage (Fig. 1c), we designed an array of cross-reactive chemiresistors in which each sensor is widely responsive to a variety of the skin-based TB VOCs.[41] The chemiresistors were based on organically stabilized spherical gold nanoparticles (core diameter 3-4 nm), 2D random networks of single-walled carbon nanotubes (RN-SWCNTs) capped with different organic layers, and polymeric composites (Fig. 3a). The response and reproducibility of the sensors (within batch and between batches) were evaluated upon exposure to octane at different concentrations (Fig. 3b). The shelf-life of the sensors was evaluated by exposing them (*e.g.* dodecanthiol-capped GNPs) to 1-methyl naphthalene at 272 ppb in nitrogen, before and after a 9 months storage, at different storage conditions: vacuum (~300 mbar), nitrogen (99.9998% pure) and room air: 18°C and 40% RH. Fig. 3c demonstrates the change in the sensors response, reflected by changes in the electrical resistivity (*i.e.* $\Delta R_{end}/R_b$), for the different storage conditions. As seen in the figure, storage is ambient conditions (room air) exhibit the largest change (35%), while the storage in vacuum and nitrogen conditions exhibit the smallest changes (19% and 17%, respectively). Nevertheless, the variances between the sensors' reading became larger in room air and nitrogen storage after 9 months of storage, in comparison to vacuum condition. Therefore, vacuum conditions were selected for storage purposes.

Prior to exposure to headspace samples of skin, the response of the sensors towards toluene was examined as a representative TB skin-based marker (Fig. 3d-e). The response of sensors was rapid, fully reversible and responsive to a wide variety of toluene concentrations, both ones that are below and above levels that are exerted from skin. As seen from the left side of Fig. 3d, different responses (positive and negative) occur upon exposing various sensors towards toluene at 0.6 ppb concentration in nitrogen. Sensors' response was evaluated also at different temperatures. As seen in the representative example in the right side of Fig. 3d, no differences were observed between the responses at different temperatures. Similar characteristics were obtained when the sensors were exposed to real skin headspace of healthy and TB subjects from both South Africa and India, each of which has its own specific combination of confounding factors (Fig. 3f).

Next, skin samples from both South Africa and India study populations were analyzed and used to train an artificial neural network (ANN), specifically, a supervised multilayer perceptron, in order to separate confirmed pulmonary active TB patients from controls. The analysis was based on arm sampling and included in total 461 samples, of which 172 were from confirmed pulmonary active TB patients and 289 from non-TB and healthy volunteers. The evaluation of the performance of the ANN was based on a *k*-fold cross-validation (*k*=6), where the global database was randomly divided into 6 parts equal in size to assess every sample. The results yielded 90.1% global accuracy, 89.5% sensitivity and 92.0% specificity after training with 46 sensors' features, from 20 sensors, identified by feature selection and 3 hidden layers, which were defined after heuristic optimization (Fig. 4a). The area under the curve of the receiver-operating curve (ROC) scored 0.9135 (Fig. 4b). By shifting slightly, the mathematical threshold (by a single point), the sensitivity of the model was increased to 90.1%, leading to a specificity of 91.3% and retaining the same accuracy. Such a change is essential for the requirements of a triage test according to the World Health Organization (WHO), in order to have score sensitivity of >90%[42]. Based on the same ANN model, a further analysis targeting sub-populations was carried out, which included discrimination between confirmed pulmonary active TB patients, and non-TB and healthy control among (i) QuantiFERON-TB Gold (QFT) positive population and (ii) HIV negative within the positive QFT population. The discrimination between confirmed pulmonary active TB and non-TB and healthy controls among QFT positive population included 290 samples: 127 confirmed pulmonary active TB cases, and 163 non-TB and healthy controls. The

discrimination between confirmed pulmonary active TB and non-TB and healthy controls among QFT positive and HIV negative population included 244 subjects: 92 confirmed pulmonary active TB cases, and 152 non-TB and healthy controls. The results of both analyses presented similar performances with 86.96-88.19% sensitivity, 92.76-93.25% specificity, and 90.57-91.03% accuracy (Fig. 4c,d). 13 potential confounding factors and their influence on the model's results were evaluated. These included gender, HIV status, time since the last bath, smoking status, and many others. Their influence was evaluated based on the discrimination accuracy of the ANN model used to discriminate between TB statuses. As can be seen from Fig. 4e, the accuracies were approximately near 50%, namely arbitrary. Thus, no significant difference within each confounding factor was found. In order to validate the potential of skin sampling, we developed an additional analysis based on Discriminant Function Analysis (DFA). Analysis of the blinded-test group (30%) resulted in 87.35 % specificity, 96.15% sensitivity and 90.65% accuracy in discriminating between active pulmonary TB patients and controls. Further details can be found in section 4 in the SI (Figure S7 and Table S4).

As an additional step for realization towards a point of care diagnostic tool ,an online exploratory pilot study was conducted by utilizing a wearable electronic device directly on the skin (Fig. 1d). The study was conducted with a study cohort of 29 healthy subjects and 18 confirmed active pulmonary TB patients. The device was placed on the chest and anterior part of the arm and included 8 nanomaterial-based sensors array. Post-processing and a linear DFA analysis resulted in a model based on leave-one-out validation with 82.76% specificity, 100% sensitivity, and 89.36% accuracy in discriminating between active pulmonary TB patients and controls from sampling the anterior arm area (Fig. 5 and Section 5 in SI).

Discussion

Studies on the detection of TB VOCs have been previously reported from the headspace of TB cells[43–45], human exhaled breath[46–50] and urine samples[51]. However, to date the use of these techniques has been impeded by the compliance of the suspected subject, a need for moderately to highly expensive equipment, high levels of expertise required to operate the instruments, the speed required for sampling and analysis, and the requirement for pre-concentration techniques[52]. The GC-MS analysis results presented herein provide the first evidence for a TB-related VOC profile emitted into the skin headspace. Possible connections between these VOCs and TB pathogenesis, and host-cholesterol intake and degradation, are also presented. Most of the introduced VOCs related to the disease are reported for the first time. Toluene is a shared VOC in both geographical locations, with similar proportions among the tested groups, identified and quantitated and has a potential metabolic pathway related to TB. As a result, toluene can be considered to be a key biomarker to discriminate between different TB statuses.

The use of a nanomaterial-based sensor array in association with pattern recognition algorithms reveals the remarkable potential that, on one hand creates a systematic screening for active case finding, and on the other hand rules out those people who do not present active pulmonary TB with high certainty. The results obtained meets the WHO's target product profile criteria for a new TB triage test, expected to surpass 90% sensitivity and 70% specificity[42], without being affected by confounding factors, *e.g.* HIV status. VOC patterns are unique for every disease; therefore, the presence of one disease should not screen other diseases.[53] As HIV is highly linked with TB infection, this allows one to eliminate possible cross-effects in the diagnosis, as demonstrated in the global TB disease classifier. Advanced discrimination among sub-populations with both positive QFT and negative HIV

statuses drew attention to the ability in distinguishing between confirmed active pulmonary TB and latent TB, or extra-pulmonary TB disease in some rare cases. These results strengthen the potential of sensing TB-related VOCs as a suitable method also to detect and diagnose latent cases, regardless of geographical differences. The pilot study with a wearable device was a further step towards assimilation of the developed sensor-based system to be applied in real-time at healthcare facilities without the need for expensive laboratory equipment. Improvement in sensor array performance can be accomplished with the new TB-related VOC profiles from the skin, which allow further customization of the sensing layer according to the GC-MS findings. Intelligent modeling via feature selection and ANNs allows interpreting the sensor signals and aiding further TB-tailored design. Finally, implementing the sensor array approach into an adhesive bandage is an additional step towards a simple and cost-effective wearable sensing patch to establish a platform to address the risk of TB epidemic in both developing and developed countries.

Methods

Study design. In this multicentric study, absorbent skin patches for capturing the VOCs were developed at Technion, IIT, Israel. These patches were sent to the 2 collaborating centers (All India Institute of Medical Sciences (AIIMS), New Delhi, India, and Groote Schuur Hospital, Cape Town, South Africa.) for VOCs sampling from the study groups. Sample collection took place in Groote Schuur Hospital in Cape Town from August 2015 until November 2016. During April 2016 and June 2017, samples were collected at AIIMS hospital in New Delhi. All participants signed informed consent forms. The clinical trial received ethical approval by the Ethical Committee of the respective hospitals: AIIMS, New Delhi: IEC/NP-103/13.03.2015, RP-39/2015 and University of Cape Town: 307/2014. The study design included 3 groups at each site with 105 participants per group: confirmed pulmonary active TB cases, healthy volunteers and confirmed non-TB cases. The clinical classification referred to 2 gold standards: sputum culture on liquid medium [BACTEC Mycobacteria Growth Indicator Tube MGIT 960 System (MGIT 960)] and GexeXpert MTB/RIF. Moreover, all the participants were screened for HIV and QuantiFERON-TB Gold In-Tube (QFT-TB) tests for further evaluation on the effect of potential confounding factors. The participants were aged between 18-85 years and the following inclusion criteria were applied. Volunteers with a skin disease that was precluded at the sampling area were excluded from the studies. In addition, smoking within half an hour prior to testing was an additional exclusion criterion. The inclusion criteria for confirmed pulmonary active TB patients included: 1) clinical symptoms 2) positive microbiology (either a positive GeneXpert MTB/RIF or/and MGIT culture for *M. tb*; 3) newly diagnosed patients. For the non-TB patients, the inclusion criteria included 1) clinical symptoms; 2) negative culture result (for HIV infected and uninfected) or Negative GeneXpert test result (HIV uninfected only); 3) chest x-rays not supporting the diagnosis of active TB; 4) no clinical symptoms at follow-up at 8 weeks. The inclusion criterion for healthy controls were: 1) no clinical symptoms in the past 12 months. For the Indian site, all the samples were collected at a single location with the same staff. For the SA site, the sampling was done in three clinics within Cape Town city with the same staff. The study cohort was designed initially for sampling with PDMS in both body locations and Tenax in the chest area only. The decision to include Tenax sampling in the anterior arm area was made after the beginning of the sampling in the two clinical sites. Therefore, the number of Tenax-based samples from anterior arm area was lower than other sampling procedures. In addition, before statistical analysis, samples were excluded due to technical reasons such as broken vials during shipment, or during GCMS failures. For the exploratory study for the wearable device, similar inclusion and exclusion criteria were applied Riga, Latvia (Nr.12-A/19). The study included 18 confirmed pulmonary active TB patients and 29 healthy controls.

Preparation of poly(2,6-diphenylphenylene oxide)-based pouch as off-line sampling tool. 20/35 meshed poly(2,6-diphenylphenylene oxide) was used as an absorbing material. Poly(2,6-diphenylphenylene oxide) (134 mg) from Buchem BV, The Netherlands, was thermal conditioned at 300°C in a constant flow of pure nitrogen for 180 min. Poly(2,6-diphenylphenylene oxide) polymer was conditioned in a glass tube with conditioned glass wool at 240°C for 48 h. After conditioning it was stored in polyester (SAATI, PE AM 120.31 PW) meshed pouches (40.3 mm X 65.11 mm, mesh opening: 47μ), which had been cleaned with a solution of 5% Decon 90 (Decon Laboratories Limited) decontaminant in distilled water (18.2 MΩ) and later stored in a vacuum oven at 100°C for >15 h. The absorbing materials were stored in vials, closed, and wrapped with Parafilm at average temperature of 4ºC.

Sampling procedure. Each participant had to wear 2 poly(2,6-diphenylphenylene oxide) envelopes on an anterior arm area, which had to be analyzed by both GC/MS and a nanomaterial-based sensor array. In addition, room samples were collected for each participant to monitor the exogenous VOCs during skin sampling. The absorbent materials were placed on the skin after cleaning with an alcohol pad (saturated with >70% isopropyl alcohol) for 10 min before sampling. The absorbent materials were covered with aluminum foil and sealed with medical adhesive tape to avoid any VOCs absorption from the surrounding environment. No shaving procedure was done in order to avoid injure to the skin and change the VOC pattern. Sampling was done over 1h. A questionnaire was filled in for every participant and the absorbent material vial numbers were also recorded. The questionnaire included data regarding the main content of the last food and drink taken prior to sampling, hygiene, vaccinations, genetic diseases, chronic diseases, infectious diseases, family TB history, smoking and drinking habits, allergies, medications and vitamins, among other details. For the room samples, the poly(2,6-diphenylphenylene oxide) patch was placed on the table near the participant for 1h to be exposed to the room air. During the whole sampling process, participants wore face masks. After sampling, the absorbing materials were stored in vials, closed and wrapped with Parafilm. These samples were stored in a refrigerator in an average temperature of 4ºC up for maximum period time of 8 months. The air transportation of the samples was with the same conditions. Opening of the vails was done at biological hood in BSL2+ laboratory with the needed protection equipment. The disinfecting material was Oosafe® Surface Disinfectant (SparMED, Denmark) that do not contain alcohol and Bactericide (confirmed for M. tuberculosis). The manufacturer claim that these Disinfectants do not release VOCs. Prior the instrumental runs of the samples, the polyester pouches were cut, and the poly(2,6-diphenylphenylene oxide) power was transferred into a glass tube containing glass wool stopper. After the transfer, the second opening was closed manually with the glass wool as well.

Sample analysis with the Gas Chromatography–Mass Spectrometry (GC-MS). An analytical evaluation of the compounds absorbed on the poly(2,6-diphenylphenylene oxide) was done with a GC/MS-QP2010 instrument (Shimadzu Corporation). It was equipped with a SLB-5ms capillary column (with 5% phenyl methyl siloxane; 30 meters in length; 0.25 mm internal diameter; 0.5 mm thicknesses; purchased from Sigma-Aldrich), and was combined with a thermal desorption (TD) system (TD20; Shimadzu Corporation). Samples were analyzed by the GC-system in split mode (20%) at 30 cm/sec constant linear speed and in a 0.70 ml/min column flow. The following oven temperature profile was set: (a) 6 min at 40ºC; (b) 13ºC/min ramp up until 170ºC; (c) a hold-time 2 min; (d) 6ºC/min ramp up until 300ºC; and (e) 15 min at 300ºC. The run duration was 55 min in total. A mixture of alkane standard solution C_8-C_{20} (Sigma-Aldrich) in hexane solvent was used as an external standard for GC/MS system calibration and normalization of the retention times and abundance changes as a result of a column aging. Compounds present in >80%[54] of skin samples until 30 minutes of retention time due to compound release from Tenax and glass wool components. Sample chromatograms were further analyzed using an open source program OpenChrom® Community Edition, version 1.1, and custom codes using MATLAB version 9.5.0.944444

(R2018b). The chromatograms were converted into txt files with the following batch processing: 1. Denoising filter (M.Z. 73, 75, 28, 147, 207, 221, 281, 295, 335, 429); 2. Savitzky-Golay filter; 3. smoothed TIC baseline detector; 4. Peak detector first derivative (MSD); and 5. Combined integrator trapezoid. These analyses steps were programed in order to overcome retention time shifts due to a prolong study run and changes in the detector sensitivity. The statistical evaluation was based on adjusted p-value for multiple peaks evaluation using a non-parametric Kruskal-Wallis test and a Steel method in comparison to the TB group as a post hoc testing. Non-parametric tests as well as the ROC curve derived Youden's cut-off point were run with JMP, version 14.0.0 (SAS Institute Inc., Cary, NC, USA, 1989–2005). For ethyl-cyclopropane analysis, 60 samples were excluded due to a prior close by saturated peak of IPA, leading to the following tested groups: (i) 85 confirmed pulmonary active TB patients; (ii) 182 non-TB patients with healthy controls; and (iii) 148 room samples.

Calibration curves of VOCs for GCMS. Identification and quantification of the VOCs that were found, involved the creation of a calibration curve for each candidate. The VOCs at different concentrations were generated using a commercial permeation/diffusion tube dilution (PDTD) system (Umwelttechnik MCZ, Germany). The system allows controlling the concentration of the VOCs. Purified dry nitrogen (99.9999%) from a commercial nitrogen generator (N-30, On Site Gas Systems, USA) equipped with a nitrogen purifier was used as a carrier gas. Samples were actively absorbed on poly(2,6-diphenylphenylene oxide) tubes at the same weight (134 mg) as used for the skin sampling from the PDTD system by pumping for 2.5 min at a flow-rate of 0.2 L/min. 3-5 repetitions were done per each concentration. The following concentration were generated: Toluene: 60,100,212,300,400,583,744 ppb; Acetic acid: 700, 900, 1100, 1300 ppb and for 2-ethyl-1hexanol: 6, 20, 40, 60, 80 ppb. The samples were analyzed by the same GC/MS methods, and a calibration curve was generated and compared to the abundance range of the clinical and room samples, using a weighted linear regression with errors in abundance.

Sample analysis with the nanomaterial-based sensor array. A stainless-steel cell for exposure contained an array of 40 nanomaterial-based sensors mounted on a customized polytetrafluoroethene circuit. The sensors included gold-nanoparticles (organically-stabilized spherical Au nanoparticles (core diameter: 3-4 nm), 2D random networks of single-walled carbon nanotubes (RN-SWCNTs), and polymers capped with different organic layers. For the ANN modeling, the following sensors proved to be key: (i) Au nanoparticles covered with octadecanethiol, decanethiol, tert-dodecanethiol, butanethiol, 2-ethylhexanethiol, dibutyl disulfide, 2-nitro-4-(trifluoromethyl) benzenethiol, benzylmercaptan, 4-chlorobenzenemethanethiol, 3-ethoxythiophenol and octadecylamine. (ii) Random networks (RNs) of carbon nanotubes (CNTs) with crystal hexa-perihexabenzocoronene (HBC) with C12 chemiresistor (HBC-C12). (iii) Polymer composites black carbon with poly(propylene-urethaneureaphenyl-disulfide) PPUU-2S chemiresistor and a composite of black carbon with poly(propylene-urethaneureaphenyl-disulfide) PPUU-2S mixed with poly(urethane-carboxyphenyl-disulfide) PUC-2S chemiresistor. Details regarding the fabrication and modification of the abovementioned sensors can be found in the literature[55–58]. The poly(2,6-diphenylphenylene oxide) samples were transferred prior to analysis into an empty thermal desorption (TD) tubes (Sigma-Aldrich), containing a glass wool stoppers, compatible with the TD system. The samples were thermally desorbed at 270°C for 10 min in an auto-sampler desorption system (TD20; Shimadzu Corporation, Japan). The sample was injected into the GC-system (Shimadzu Corporation, Japan) in a direct (splitless) mode at a constant 3 mL/min total flow and the desorbed sample was temporarily stored in a stainless steel column (150°C). The samples from the TD were then delivered by a 6-way Valco valve, equipped with 10 mL stainless steel loop (VICI, Valco Instruments Company Inc., USA) into a stainless-steel chamber containing the sensors with a volume of 330 cm^3. When a one-way valve connecting the chamber to the column was open, the sample was sucked into the chamber, while the remaining volume was filled with N_2 until reaching atmospheric pressure in the chamber. A Keithley 2701 DMM

data acquisition/data-logging system was used to measure the resistance of all the sensors simultaneously as a function of time. The sensors' baseline responses were recorded for 5 min in vacuum (~30 mtorr), 5 min under pure nitrogen (99.999%), 5 min in vacuum, and 5 min under sample exposure, followed by a further 3 min under vacuum conditions. To supervise the sensor's functionality during the experiment, and to overcome possible sensor response drift, a fixed calibration gas mixture containing 11.5 ppm isopropyl alcohol, 2.8 ppm trimethylbenzene and 0.6 ppm 2-ethylhexanol was exposed to the sensors daily. This calibration gas was generated PDTD system. The calibration mixture was absorbed on a clean tube for 2 min. Several features are extracted from each of the sensor's signals upon exposure, including area under the curve, delta R peak, delta R middle and delta R end. The last 3 features are based on the difference between the baseline resistance, usually during vacuum, and the resistance during the response towards the exposure: peak point, middle part and the end part of the signal.

Artificial neural networks. To validate a multilayer perceptron (MLP) using *k*-fold cross-validation, the verification dataset was used to test performance. This dataset is not involved in the weight modification process, which enables this validation method. To carry it out, the global database containing samples from India and South Africa was divided randomly into *k* parts equal in size. The *k* selected was 6, and, therefore the MLP was evaluated 6 times by swapping the verification dataset for a new one in each test. In essence, for every test ~83% of training samples and ~17% of verification samples were used, and the final statistical performance of the model was evaluated by averaging the results from all *k* tests, which covered every sample in the initial database. When a model provides accurate results during this validation method, it typically signifies that it can generalize well; and therefore, it is reliable for data that is different from the one used in the training or verification datasets[59,60]. The model utilized 20 sensors with a total of 46 features. Feature selection and MLP-related calculations relied on MATLAB version 9.3.0.713579 (R2017b).

Discriminant function analysis (DFA) DFA is a statistical method for data analysis when the groups to be discriminated are defined (labeled) before analysis[61]. The input variables are the features extracted from sensors' responses towards the skin samples. The decision on either linear or quadratic model was based on homogeneity of the variance-covariance matrices of the tested groups according to statistical tests, e.g. Bartlett's[62,63]. During this study equal prior probabilities were set to confirmed pulmonary active TB and non-TB with healthy volunteers, respectively. For the off-line approach, the model was evaluated by randomly splitting the original database into 70% training set and 30% test set. For the on-line device system, a linear DFA model was applied based on three sensor features with leave-one-out validation.

Wearable device. The device was connected to the PC and the developed software was launched. Initially, the device samples room air for 10 min. Then the patch was attached to the participant. The patch was attached for 60 min. There was no direct contact between the skin and the sensors. The device included eight sensors as described in previous method sections. On exposure, several features, such as the Area under curve, Delta R peak, Delta R middle and Delta R end, are extracted from the sensor signal. The last three characteristics are based on the difference between the resistance of the baseline, typically during the room air exposure, and the resistance in the exposure response: peak point, middle and end part, respectively. The obtained a linear DFA model was based on three features from three sensors.

Data availability

The authors declare that the data supporting the findings of this study are available within the paper and its supplementary information file. Any additional data if needed will be provided upon reasonable request.

Code availability

Custom codes are available on request.

References

1. Lalvani, A. & Pareek, M. A 100 year update on diagnosis of tuberculosis infection. *Br. Med. Bull.* **93**, 69–84 (2010).
2. Russell, D. G. Mycobacterium tuberculosis: Here today, and here tomorrow. *Nature Reviews Molecular Cell Biology* **2**, 569–577 (2001).
3. Williams, V. G. Tuberculosis: clinical features, diagnosis and management. *Nurs. Stand.* **20**, 49–53 (2006).
4. de Vries, S. G. *et al.* Barriers and facilitators to the uptake of tuberculosis diagnostic and treatment services by hard-to-reach populations in countries of low and medium tuberculosis incidence: a systematic review of qualitative literature. *The Lancet Infectious Diseases* **17**, e128–e143 (2017).
5. World Health Organization. *Global tuberculosis report 2018*. (2018).
6. Lu, C. *et al.* A systematic review of reported cost for smear and culture tests during multidrug-resistant Tuberculosis treatment. *PLoS One* **8**, 1–14 (2013).
7. Siddiqi, K., Lambert, M. L. & Walley, J. Clinical diagnosis of smear-negative pulmonary tuberculosis in low-income countries: The current evidence. *Lancet Infectious Diseases* **3**, 288–296 (2003).
8. World Health Organization. *WHO | The global burden of disease: 2004 update*. *WHO* (World Health Organization, 2014).
9. *Toman's tuberculosis case detection, treatment and monitoring*. (World health organization, 2004).
10. Steingart, K. R. *et al.* Xpert® MTB/RIF assay for pulmonary tuberculosis and rifampicin resistance in adults. *Cochrane Database Syst. Rev.* 1–168 (2014). doi:10.1002/14651858.CD009593.pub3
11. Pathmanathan, I. *et al.* Rolling out Xpert MTB/RIF® for tuberculosis detection in HIV-positive populations: An opportunity for systems strengthening. *Afr. J. Lab. Med.* **6**, 1–9 (2017).
12. *High-priority target product profi les for new tuberculosis diagnostics: report of a consensus meeting*.
13. http://www.who.int/trade/distance_learning/gpgh/gpgh3/en/index7.html.
14. Riazanskaia, S., Blackburn, G., Harker, M., Taylor, D. & Thomas, C. L. The analytical utility of thermally desorbed polydimethylsilicone membranes for in-vivo sampling of volatile organic compounds in and on human skin. *Analyst* **133**, 1020–1027 (2008).
15. Frank Kneepkens, C. M., Lepage, G. & Roy, C. C. The potential of the hydrocarbon breath test as a measure of lipid peroxidation. **17**, 127–160 (1994).
16. Lechner, M. & Rieder, J. Mass spectrometric profiling of low-molecular-weight volatile compounds-Diagnostic potential and latest applications. *Curr. Med. Chem.* **14**, 987–995 (2007).
17. Abaffy, T. *et al.* Differential Volatile Signatures from Skin, Naevi and Melanoma: A Novel Approach to Detect a Pathological Process. *PLoS One* **5**, e13813 (2010).

18. Turner, C. Potential of breath and skin analysis for monitoring blood glucose concentration in diabetes. *Exp. Rev. Mol. Diag.* **11**, 497–503 (2011).
19. Cortesia, C. *et al.* Acetic Acid, the active component of vinegar, is an effective tuberculocidal disinfectant. *MBio* **5**, 1–4 (2014).
20. Kolk, A. H. J. *et al.* Breath-based biomarkers for Tuberculosis. in **8371**, 83710A–10 (International Society for Optics and Photonics, 2012).
21. de Lacy Costello, B. *et al.* A review of the volatiles from the healthy human body. *J. Breath Res.* **8**, 1–29 (2014).
22. Api, A. M. *et al.* RIFM fragrance ingredient safety assessment, hexyl butyrate, CAS Registry Number 2639-63-6. *Food and Chemical Toxicology* **130**, 110608 (2019).
23. Hexyl butyrate - Substance Information - ECHA. Available at: https://echa.europa.eu/substance-information/-/substanceinfo/100.018.306. (Accessed: 12th October 2020)
24. Tay, S. T. *et al.* Two new Mycobacterium strains and their role in toluene degradation in a contaminated stream. *Appl. Environ. Microbiol.* **64**, 1715–20 (1998).
25. Wichmann, G. *et al.* An experimental model for the determination of immunomodulating effects by volatile compounds. *Toxicol. Vitr.* **19**, 685–693 (2005).
26. Harris, J. *et al.* T helper 2 cytokines inhibit autophagic control of intracellular Mycobacterium Tuberculosis. *Immunity* **27**, 505–517 (2007).
27. Kim, H. *et al.* Cytochrome P450 isozymes responsible for the metabolism of toluene and styrene in human liver microsomes. *Xenobiotica* **27**, 657–665 (1997).
28. Tassaneeyakul, W. *et al.* Human cytochrome P450 isoform specificity in the regioselective metabolism of toluene and o-, m- and p-xylene. *J. Pharmacol. Exp. Ther.* **276**, 101–8 (1996).
29. Klenø, J. G., Wolkoff, P., Clausen, P. A., Wilkins, C. K. & Pedersen, T. Degradation of the adsorbent tenax TA by nitrogen oxides, ozone, hydrogen peroxide, OH radical, and limonene oxidation products. *Environ. Sci. Technol.* **36**, 4121–4126 (2002).
30. Clausen, P. A. & Wolkoff, P. Degradation products of Tenax TA formed during sampling and thermal desorption analysis: Indicators of reactive species indoors. *Atmos. Environ.* **31**, 715–725 (1997).
31. Chu, L., Deng, S., Zhao, R., Deng, J. & Kang, X. Comparison of Adsorption/Desorption of Volatile Organic Compounds (VOCs) on Electrospun Nanofibers with Tenax TA for Potential Application in Sampling. *PLoS One* **11**, e0163388 (2016).
32. Brown, V. M., Crump, D. R., Plant, N. T. & Pengelly, I. Evaluation of the stability of a mixture of volatile organic compounds on sorbents for the determination of emissions from indoor materials and products using thermal desorption/gas chromatography/mass spectrometry. *J. Chromatogr. A* **1350**, 1–9 (2014).
33. MacLeod, G. & Ames, J. M. Comparative assessment of the artefact background on thermal desorption of tenax GC and tenax TA. *J. Chromatogr. A* **355**, 393–398 (1986).
34. Hwan Lee, J., Batterman, S. A., Jia, C. & Chernyak, S. Ozone artifacts and carbonyl measurements using tenax GR, tenax TA, carbopack B, and carbopack X adsorbents. *J. Air Waste Manag. Assoc.* **56**, 1503–1517 (2006).
35. Cao, X. L. & Hewitt, C. N. Study of the Degradation by Ozone of Adsorbents and of Hydrocarbons Adsorbed during the Passive Sampling of Air. *Environ. Sci. Technol.* **28**, 757–762 (1994).

36. Barkan, D., Hedhli, D., Yan, H.-G., Huygen, K. & Glickman, M. S. Mycobacterium tuberculosis lacking all mycolic acid cyclopropanation is viable but highly attenuated and hyperinflammatory in mice. *Infect. Immun.* **80**, 1958–68 (2012).
37. Daffé, M., Quémard, A. & Marrakchi, H. Mycolic acids: From chemistry to biology. in *Biogenesis of Fatty Acids, Lipids and Membranes* (ed. Geiger, O.) 182–206 (Springer, Cham, 2019). doi:10.1007/978-3-319-50430-8_18
38. Chang, S. W. *et al.* Gut Hormones, Appetite Suppression and Cachexia in Patients with Pulmonary TB. *PLoS One* **8**, e54564 (2013).
39. Yurt, S. *et al.* The role of feed regulating peptides on weight loss in patients with pulmonary tuberculosis. *Clin. Biochem.* **46**, 40–44 (2013).
40. Salthammer, T. Emissions of Volatile Organic Compounds from Products and Materials in Indoor Environments. *Handb. Environ. Chem.* **4**, 37–71 (2004).
41. Röck, F., Barsan, N. & Weimar, U. Electronic nose: Current status and future trends. *Chem. Rev.* **108**, 705–725 (2008).
42. World Health Organization. *WHO | High-priority target product profiles for new tuberculosis diagnostics*. *World Health Organization* (World Health Organization, 2014).
43. Syhre, M. & Chambers, S. T. The scent of Mycobacterium tuberculosis. *Tuberculosis* **88**, 317–323 (2008).
44. Nawrath, T., Mgode, G. F., Weetjens, B., Kaufmann, S. H. & Schulz, S. The volatiles of pathogenic and nonpathogenic mycobacteria and related bacteria. *Beilstein J. Org. Chem.* **8**, 290–299 (2012).
45. Mgode, G. F. *et al.* Ability of Cricetomys rats to detect Mycobacterium tuberculosis and discriminate it from other microorganisms. *Tuberculosis* **92**, 182–186 (2012).
46. Phillips, M. *et al.* Breath biomarkers of active pulmonary tuberculosis. *Tuberculosis* **90**, 145–151 (2010).
47. Phillips, M. *et al.* Point-of-care breath test for biomarkers of active pulmonary tuberculosis. *Tuberculosis* **92**, 314–320 (2012).
48. Phillips, M. *et al.* Volatile biomarkers of pulmonary tuberculosis in the breath. *Tuberculosis* **87**, 44–52 (2007).
49. Nakhleh, M. *et al.* Detecting active pulmonary tuberculosis with a breath test using nanomaterial-based sensors. *Eur. Respir. J.* **43**, 1522–1525 (2014).
50. Syhre, M., Manning, L., Phuanukoonnon, S., Harino, P. & Chambers, S. T. The scent of Mycobacterium tuberculosis–Part II breath. *Tuberculosis* **89**, 263–266 (2009).
51. Banday, K. M. *et al.* Use of urine volatile organic compounds to discriminate tuberculosis patients from healthy subjects. *Anal. Chem.* **83**, 145526–5534 (2011).
52. Miekisch, W., Schubert, J. & Noeldge-Schomburg, G. Diagnostic potential of breath analysis—focus on volatile organic compounds. *Clin. Chim. Acta* **347**, 25–39 (2004).
53. Nakhleh, M. K. *et al.* Diagnosis and Classification of 17 Diseases from 1404 Subjects via Pattern Analysis of Exhaled Molecules. *ACS Nano* **11**, 112–125 (2017).
54. Ruiying, C. *et al.* A comprehensive analysis of metabolomics and transcriptomics in non-small cell lung cancer. *PLoS One* **15**, e0232272 (2020).
55. Peng, G. *et al.* Diagnosing lung cancer in exhaled breath using gold nanoparticles. *Nat. Nanotechnol.* **4**, 669–673 (2009).
56. Dovgolevsky, E., Tisch, U. & Haick, H. Chemically sensitive resistors based on monolayercapped cubic nanoparticles: Towards configurable nanoporous sensors. *Small* **5**, 1158–1161 (2009).

57. Zilberman, Y., Ionescu, R., Feng, X., Müllen, K. & Haick, H. Nanoarray of polycyclic aromatic hydrocarbons and carbon nanotubes for accurate and predictive detection in real-world environmental humidity. *ACS Nano* **5**, 6743–6753 (2011).
58. Huynh, T.-P. *et al.* Composites of Polymer and Carbon Nanostructures for Self-Healing Chemical Sensors. *Adv. Mater. Technol.* **1**, (2016).
59. Cancilla, J. C., Díaz-Rodríguez, P., Izquierdo, J. G., Bañares, L. & Torrecilla, J. S. Artificial neural networks applied to fluorescence studies for accurate determination of N-butylpyridinium chloride concentration in aqueous solution. *Sens. Actuator B-Chem.* **198**, 173–179 (2014).
60. Soleymani, A. R., Saien, J. & Bayat, H. Artificial neural networks developed for prediction of dye decolorization efficiency with UV/K2S2O8 process. *Chem. Eng. Sci.* **170**, 29–35 (2011).
61. Brereton, R. G. *Chemometrics : applications of mathematics and statistics to laboratory systems*. (E. Horwood, 1990).
62. Snedecor, G. W. & Cochran, W. G. *Statistical methods*. (Iowa State University Press, 1989).
63. Lim, T.-S. & Loh, W.-Y. A comparison of tests of equality of variances. *Comput Stat.* **22**, 287–301 (1996).

Declarations

Acknowledgements

This work received funding from the Phase-II Grand Challenges Explorations award of the Bill and Melinda Gates Foundation (Grant ID: OPP1109493) and the and Horizon 2020 ICT grant under the A-Patch project. R.V. acknowledges the Ariane de Rothschild Foundation for a PhD scholarship. The authors acknowledge Dr. Viki Kloper for help with artwork production, Mrs. Shifaa Bdarneh and Mrs. Salam Khatib for help with the experiments. S.S and K.P. acknowledge the technical support from Ms. Priyanka Sharma, Suman Chauhan, and Syed Beenish Rufai. In addition, acknowledgments to Ananda Vishal, MD, Vijay Hadda, MD and other physicians of All India Institute of Medical Sciences (AIIMS) for referring their patients. G.S and M.L acknowledge Mrs. Signe Daukste, Mrs. Ruta Deklava, Mrs. Vjaceslava Ruza and Mrs. Evita Biraua for conducting the Latvian trial.

Author Contributions

R.V. and H.H. devised and organized this investigation. R.V. carried out most of the experiments, analyzed data, and wrote the manuscript. B.R. carried out the wearable device expletory study. E.M. preformed calibration curves for the reported VOCs and sensor storage conditions study. A.E. and K.D. were responsible for the clinical trial in Cape Town, South Africa. P.K and S.S. were responsible for the clinical trial in New Delhi, India. A.G. preformed DFA analysis. J.C.C. and J.S.T. preformed ANN analysis. S.G and L.M. were responsible for the clinical trial in Riga, Lativa. E.M. preformed calibration curves for the reported VOCs. R.V., B.R., E.M., A.E., F.F., P.K., A.G., J.C.C., J.S.T., S.G., L.M., K.D., S.S., and H.H. discussed the data and revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at X

Correspondence and requests for materials should be addressed to H.H.

Figures

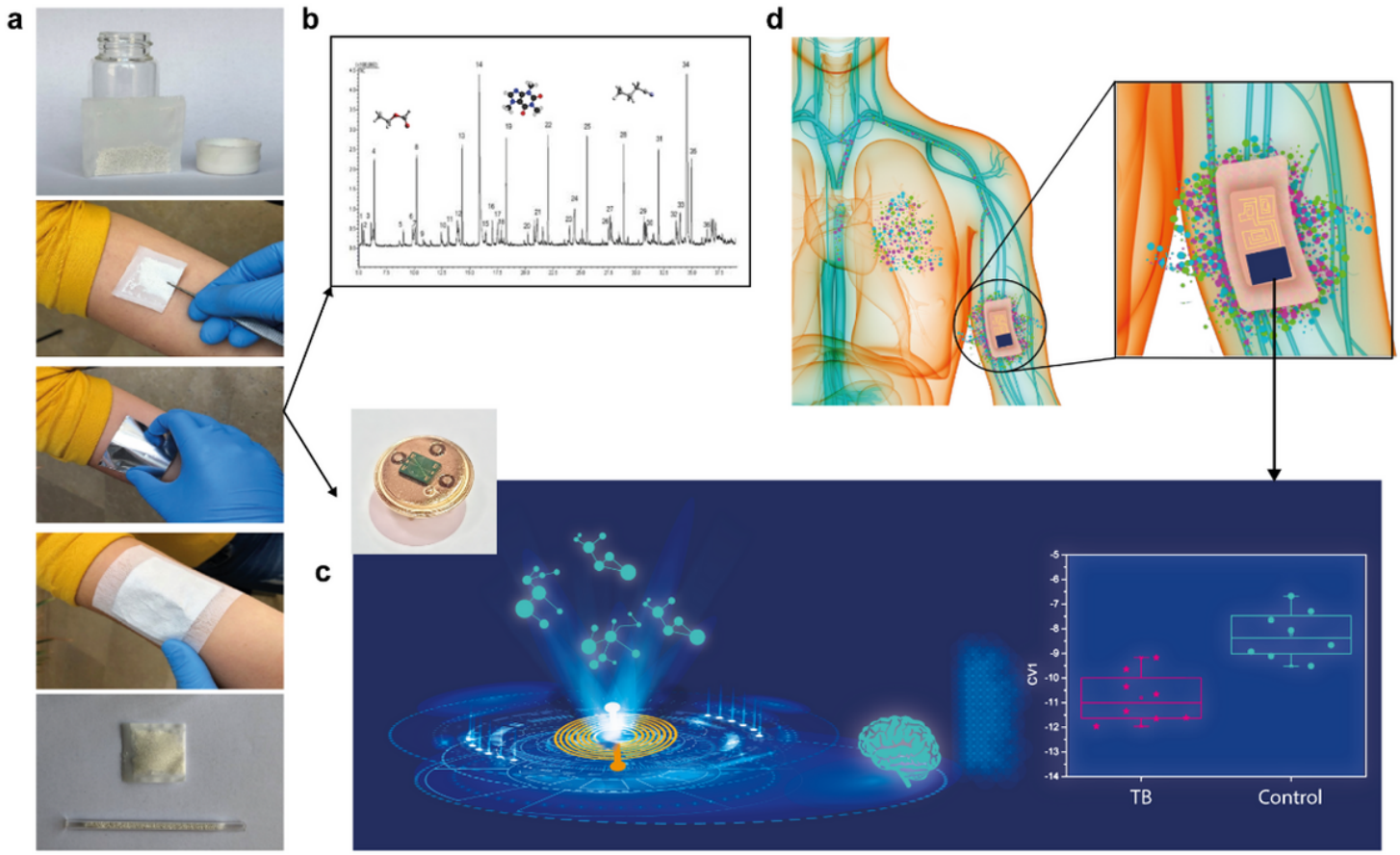

Figure 1

Study schematics. a. skin headspace sampling procedure with poly(2,6-diphenylphenylene oxide) polymer. The samples are transferred into glass tubes for two analysis: b GC-MS analysis of the collected samples; and c. nanomaterial-based sensor array in conjugation machine learning analysis of the collected samples. d. a wearable device applied directly on the skin.

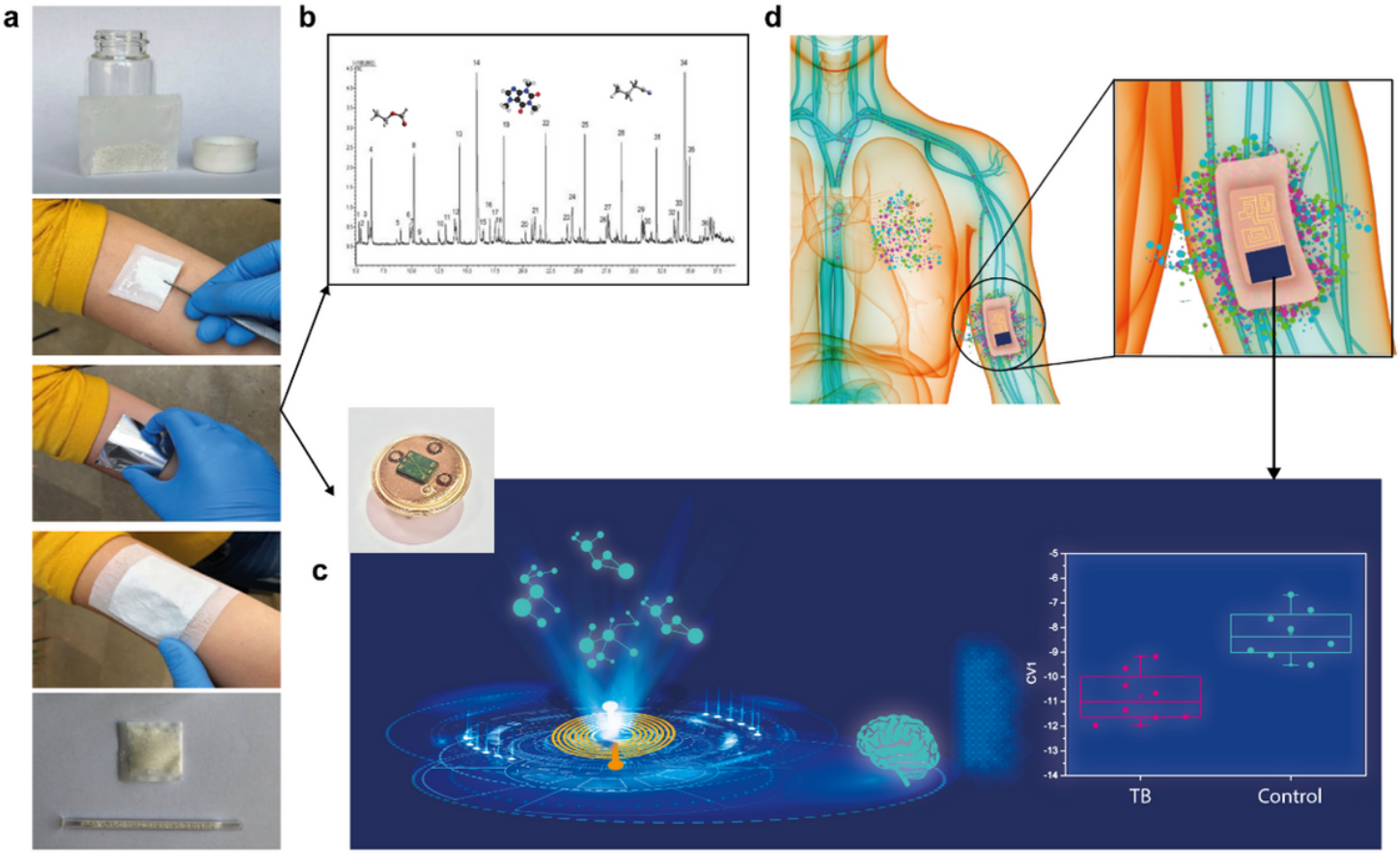

Figure 1

Study schematics. a. skin headspace sampling procedure with poly(2,6-diphenylphenylene oxide) polymer. The samples are transferred into glass tubes for two analysis: b GC-MS analysis of the collected samples; and c. nanomaterial-based sensor array in conjugation machine learning analysis of the collected samples. d. a wearable device applied directly on the skin.

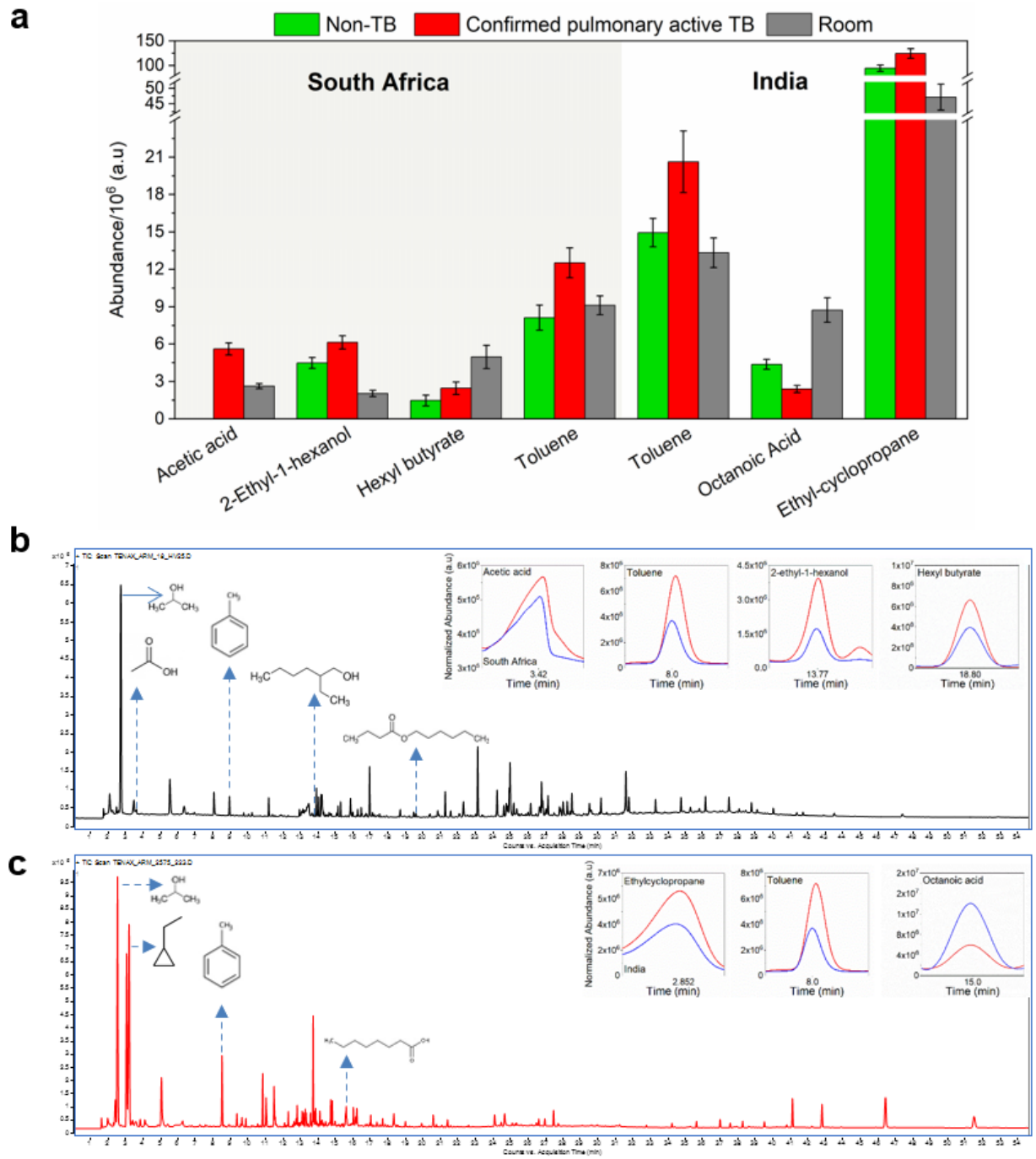

Figure 2

GC-MS results. a. An abundance of toluene, acetic acid, 2-ethyl-1-hexanol, ethyl-cyclopropane, hexyl butyrate and octanoic acid among confirmed pulmonary active TB patients, non-TB patients with healthy controls, and room samples. Error bars represent standard errors. For hexyl butyrate, two extreme outlier points were excluded for the confirmed pulmonary active TB patients. b-c Representative chromatograms with statistically significant VOCs and IPA as a skin cleaning component, from SA and India, respectively. d. Representative chromatograms of relevant VOCs based on total ion count traces.

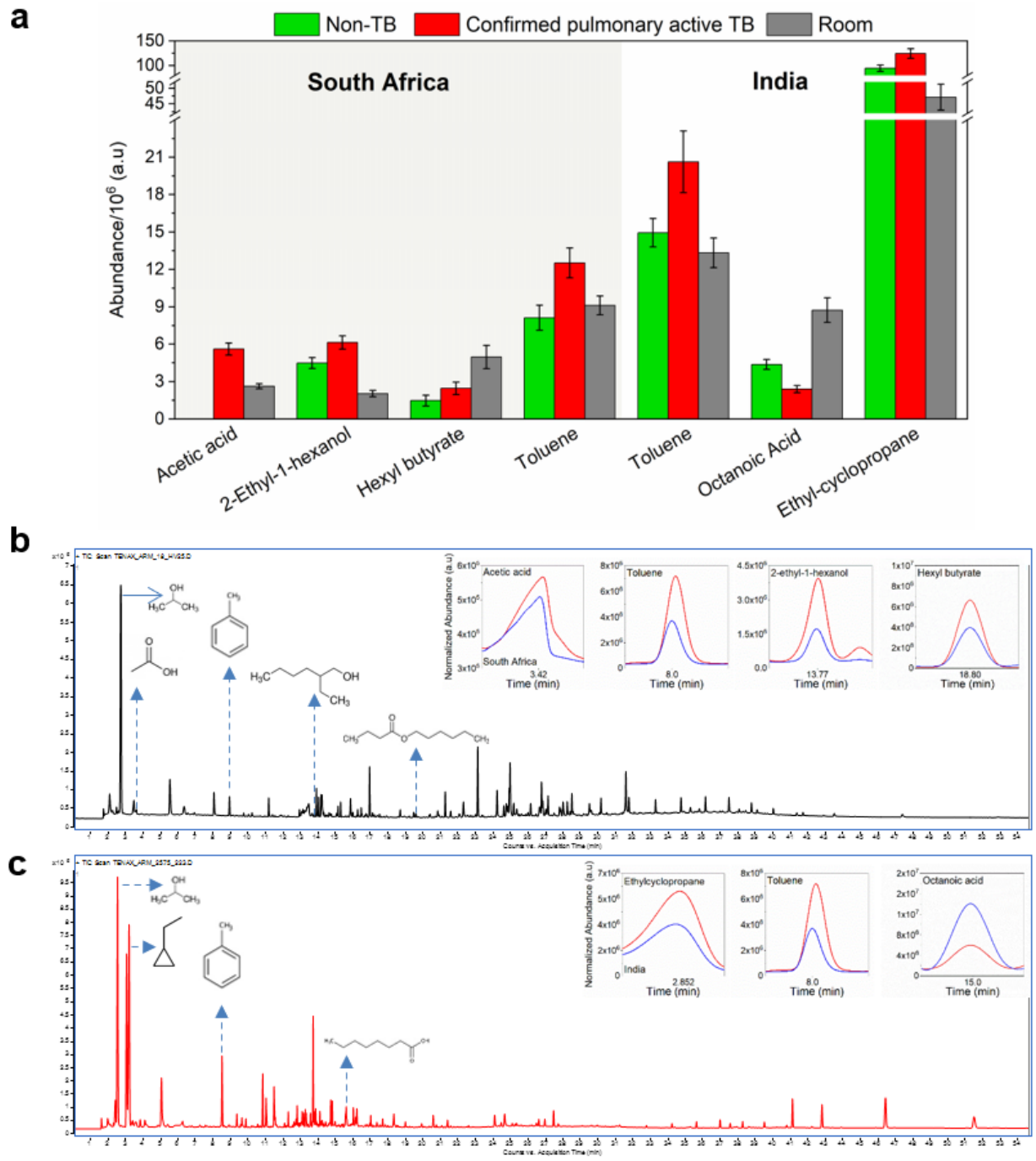

Figure 2

GC-MS results. a. An abundance of toluene, acetic acid, 2-ethyl-1-hexanol, ethyl-cyclopropane, hexyl butyrate and octanoic acid among confirmed pulmonary active TB patients, non-TB patients with healthy controls, and room samples. Error bars represent standard errors. For hexyl butyrate, two extreme outlier points were excluded for the confirmed pulmonary active TB patients. b-c Representative chromatograms with statistically significant VOCs and IPA as a skin cleaning component, from SA and India, respectively. d. Representative chromatograms of relevant VOCs based on total ion count traces.

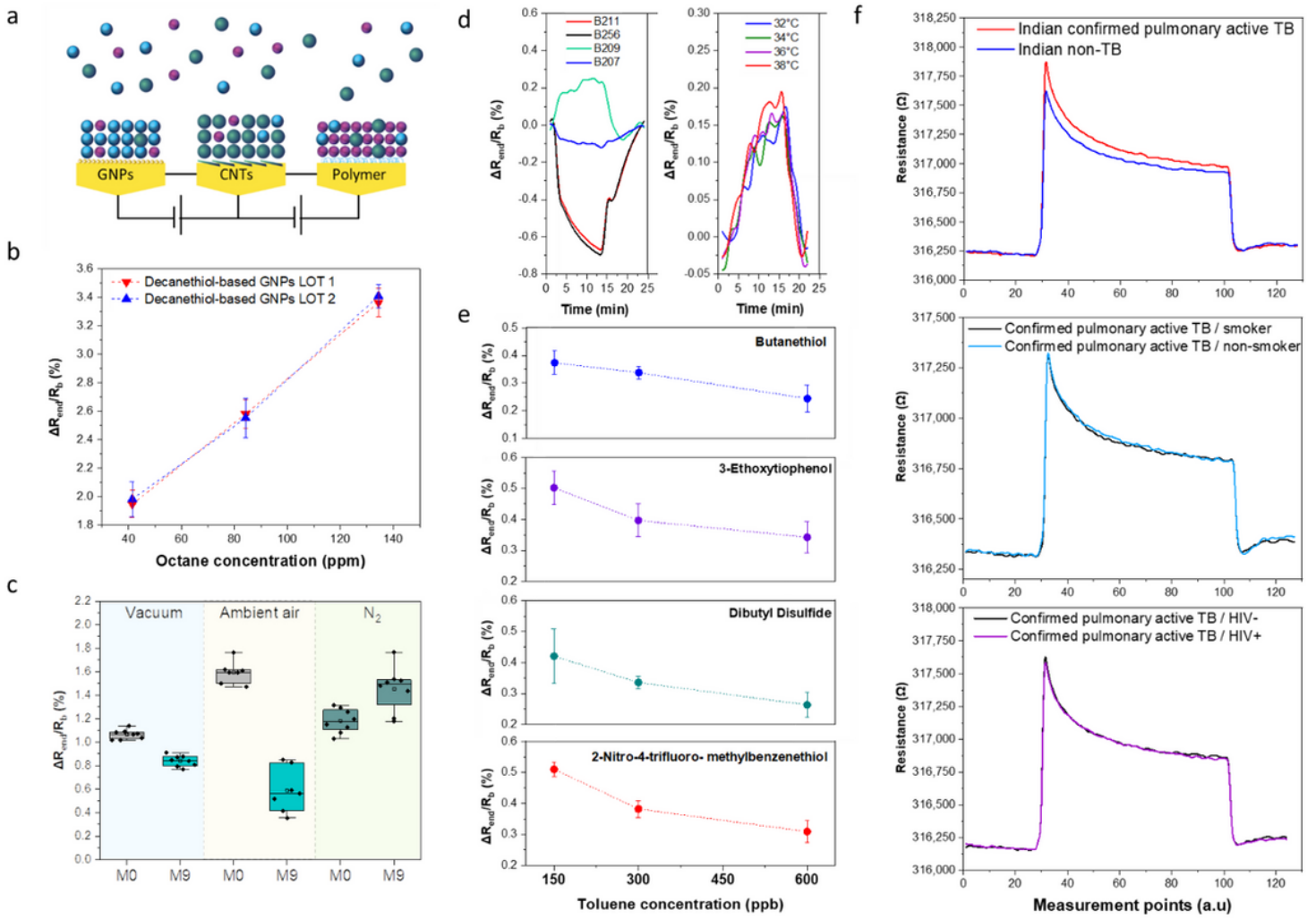

Figure 3

Sensor array responses. a. schematic illustration a sensor array. b. Representative responses of decanethiol-based GNPs sensors from two different batches towards increasing concentrations of octane. c. Representative responses of dodecanthiol-based GNPs towards 1-Methyl Naphthalene at 272 ppb in nitrogen exposure in different storage conditions at starting point (M0) and after 9 months (M9). d-e. sensors' response towards toluene exposure. d. Response rate to 0.6 ppb toluene in nitrogen of sensors based on different thiol ligands (left) and temperature effect on sensor response during decanethiol-based GNPs exposure to toluene at 1.2 ppm (right). e. Response rate of different sensors based on GNPs towards toluene range of concentration. f. Signals of the same sensor to confirmed pulmonary active TB and non-TB skin samples from Indian site, confirmed pulmonary active TB with or without smoking habits and HIV infection.

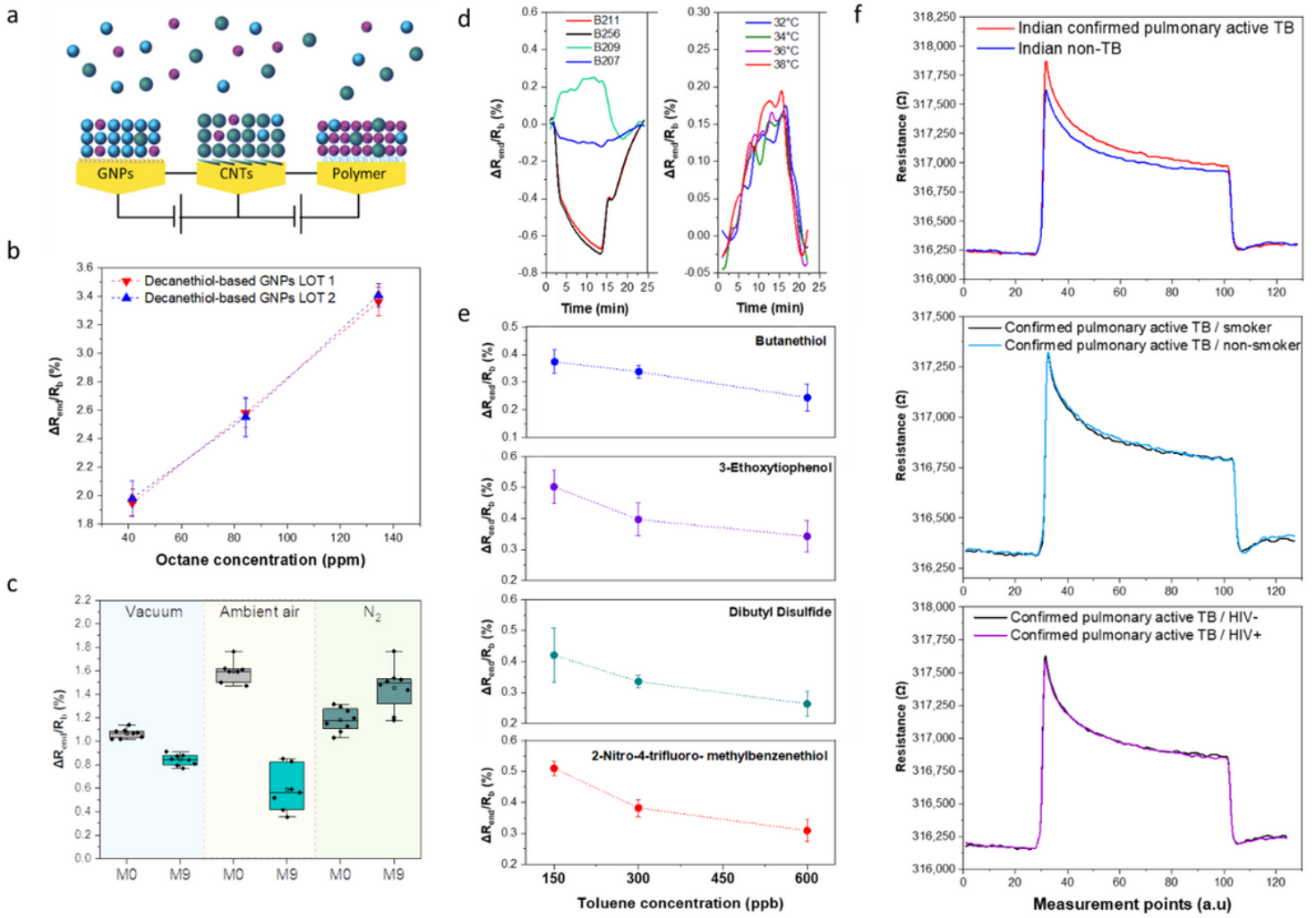

Figure 3

Sensor array responses. a. schematic illustration a sensor array. b. Representative responses of decanethiol-based GNPs sensors from two different batches towards increasing concentrations of octane. c. Representative responses of dodecanthiol-based GNPs towards 1-Methyl Naphthalene at 272 ppb in nitrogen exposure in different storage conditions at starting point (M0) and after 9 months (M9). d-e. sensors' response towards toluene exposure. d. Response rate to 0.6 ppb toluene in nitrogen of sensors based on different thiol ligands (left) and temperature effect on sensor response during decanethiol-based GNPs exposure to toluene at 1.2 ppm (right). e. Response rate of different sensors based on GNPs towards toluene range of concentration. f. Signals of the same sensor to confirmed pulmonary active TB and non-TB skin samples from Indian site, confirmed pulmonary active TB with or without smoking habits and HIV infection.

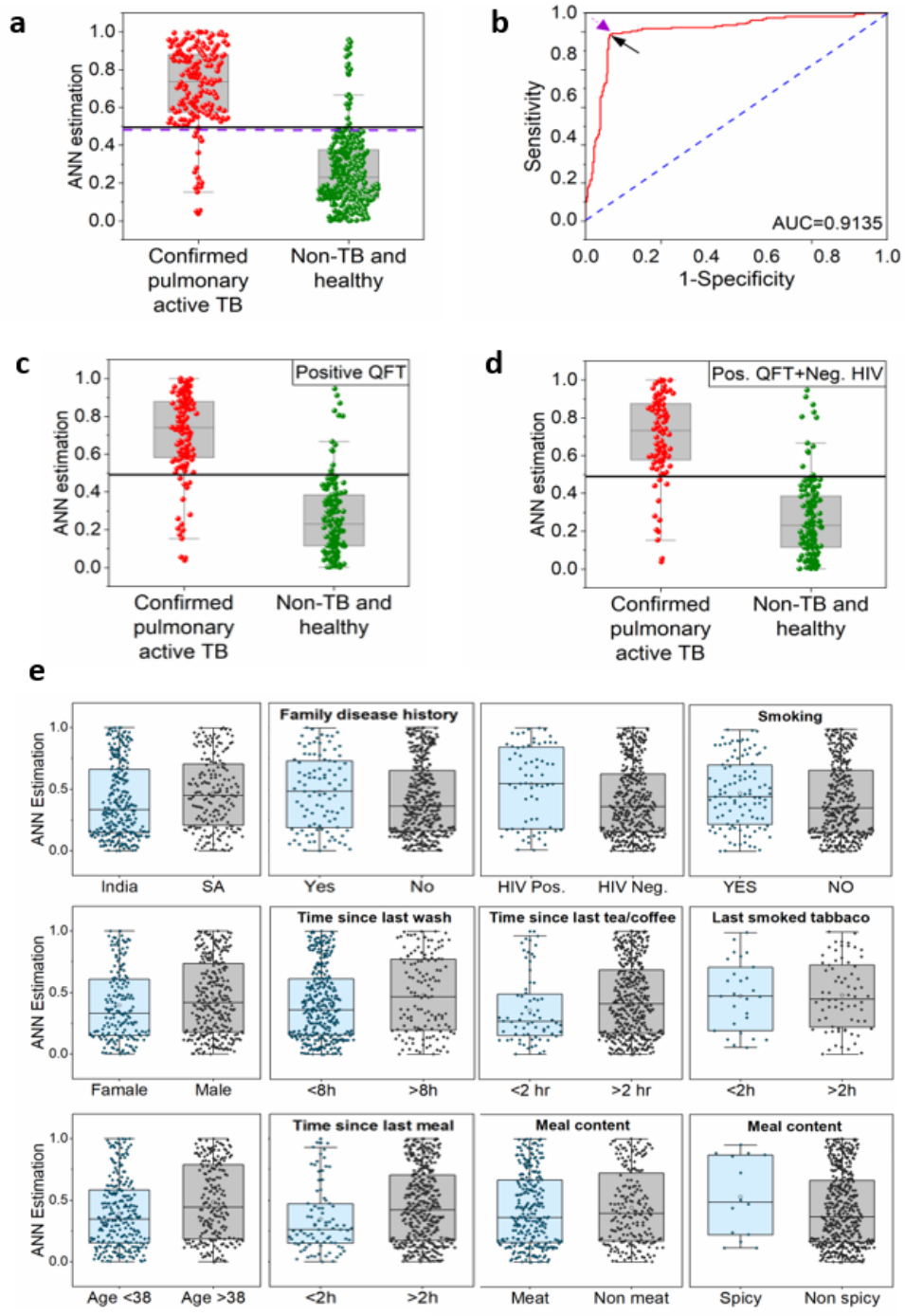

Figure 4

ANN classifier results associated confounding factors. a. Boxplot of ANN estimation score of the model. Each point represents one sample. The central solid line represents Youden's cut-off. The dashed line represents the modified cut-off threshold to meet the WHO criterion of a sensitivity >90%. Samples above the cut-off were classified as non-TB and healthy, and samples below the cut-off were classified as confirmed pulmonary active TB samples. b. Receiver operating characteristic (ROC) curve of model based on mathematically optimal Youden's cut-off (black arrow) and the modified cut-off according to WHO's guidelines (purple arrow). c. Boxplot of ANN estimation score of the model for subpopulation with QFT positive status. d. Boxplot of ANN estimation score of

the model for subpopulation with QFT positive and HIV negative statuses. e Boxplots of ANN estimation scores for confounding factors. QFT- QuantiFERON-TB Gold test.

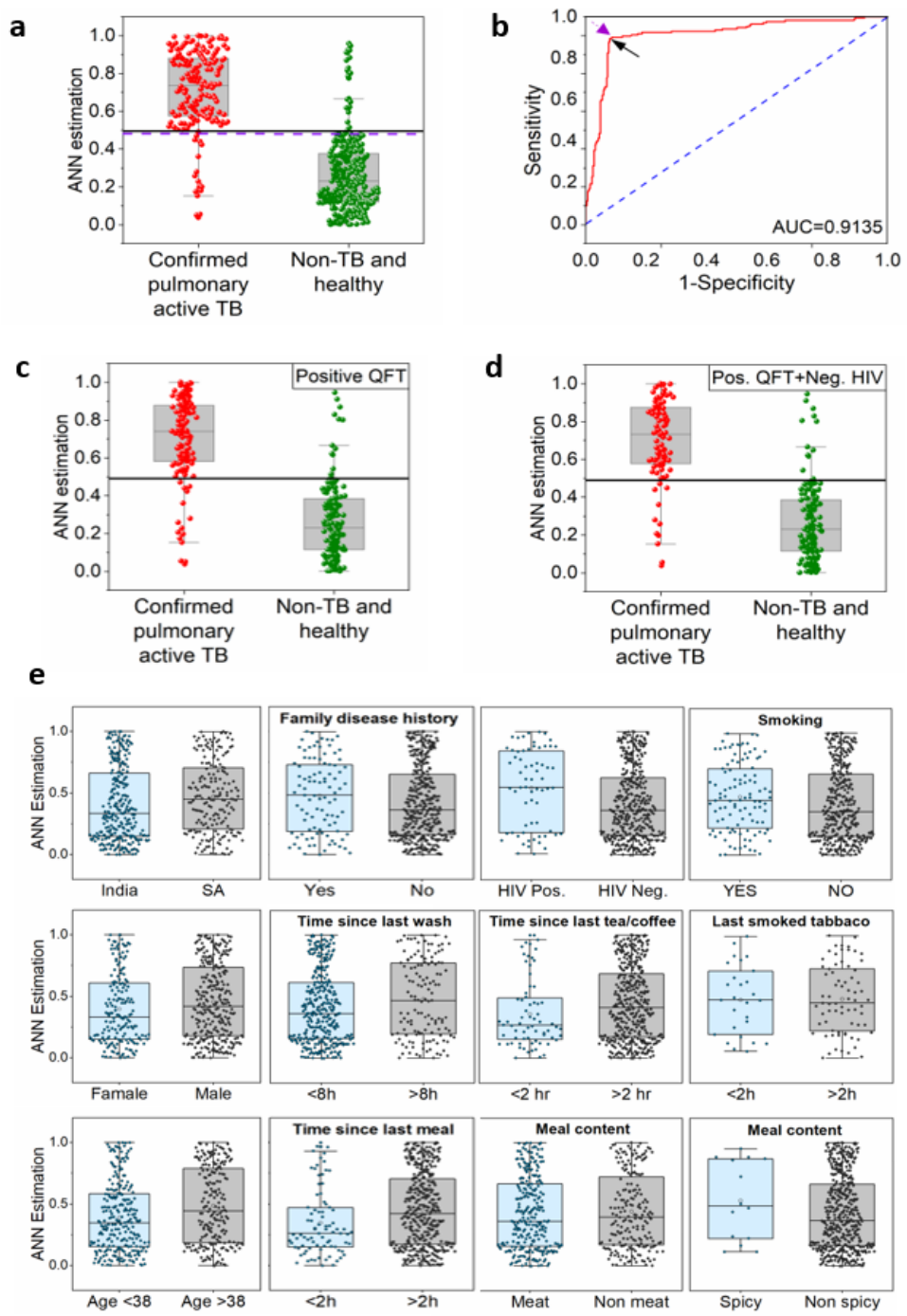

Figure 4

ANN classifier results associated confounding factors. a. Boxplot of ANN estimation score of the model. Each point represents one sample. The central solid line represents Youden's cut-off. The dashed line represents the modified cut-off threshold to meet the WHO criterion of a sensitivity >90%. Samples above the cut-off were classified as non-TB and healthy, and samples below the cut-off were classified as confirmed pulmonary active TB samples. b. Receiver operating characteristic (ROC) curve of model based on mathematically optimal Youden's cut-off (black arrow) and the modified cut-off according to WHO's guidelines (purple arrow). c. Boxplot of ANN

estimation score of the model for subpopulation with QFT positive status. d. Boxplot of ANN estimation score of the model for subpopulation with QFT positive and HIV negative statuses. e Boxplots of ANN estimation scores for confounding factors. QFT- QuantiFERON-TB Gold test.

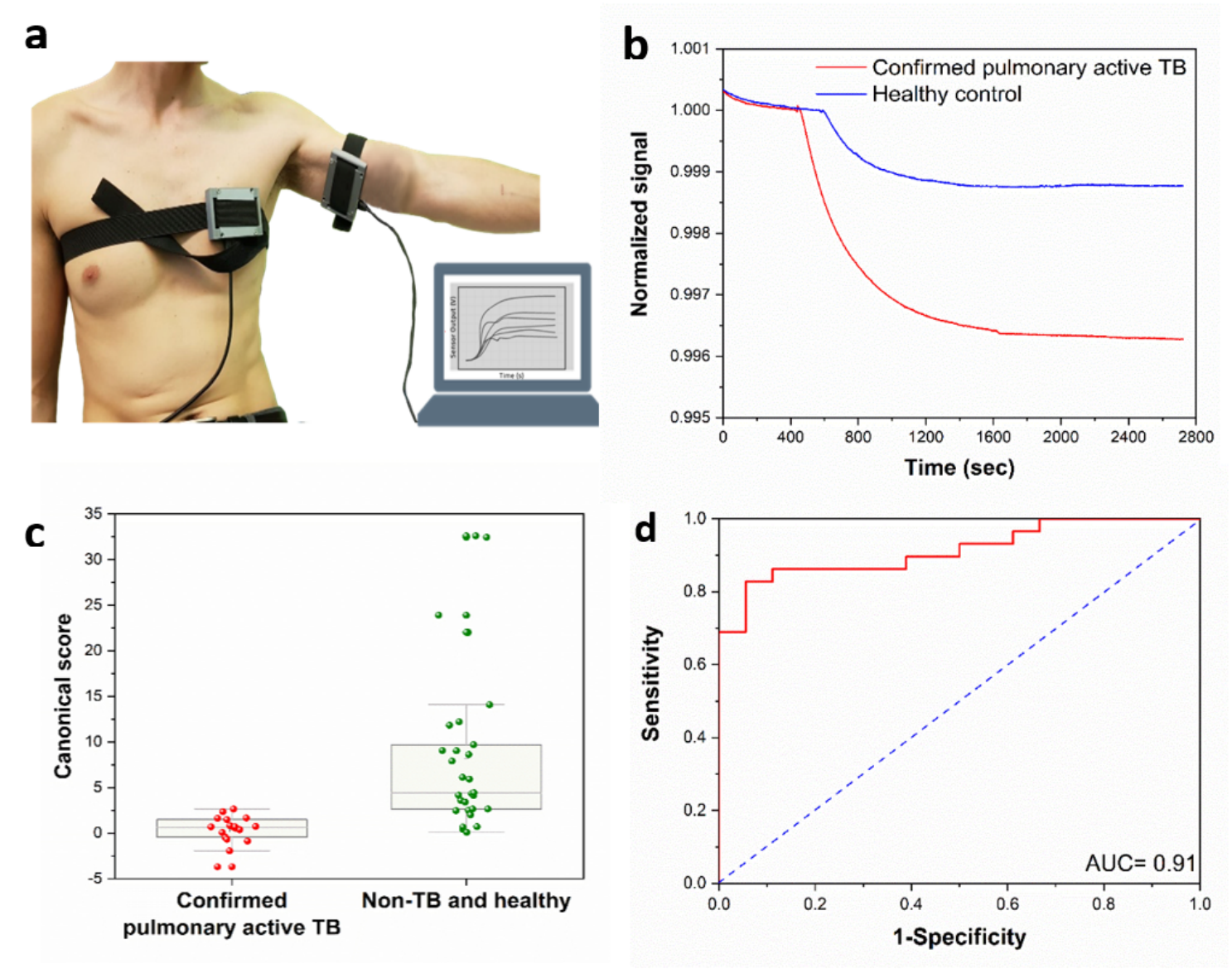

Figure 5

Wearable device TB diagnosis. a wearable devices on the chest and anterior arm of a volunteer. b representative normalized signals of one of the sensors in the wearable device attached to the anterior arm area. The plotted signal is the normalized resistance to the baseline resistance before patch is attached to the experiment participant. c boxplot of the canonical score of linear DFA model. Each point represents one sample. d ROC curve of model. AUC= area under curve.

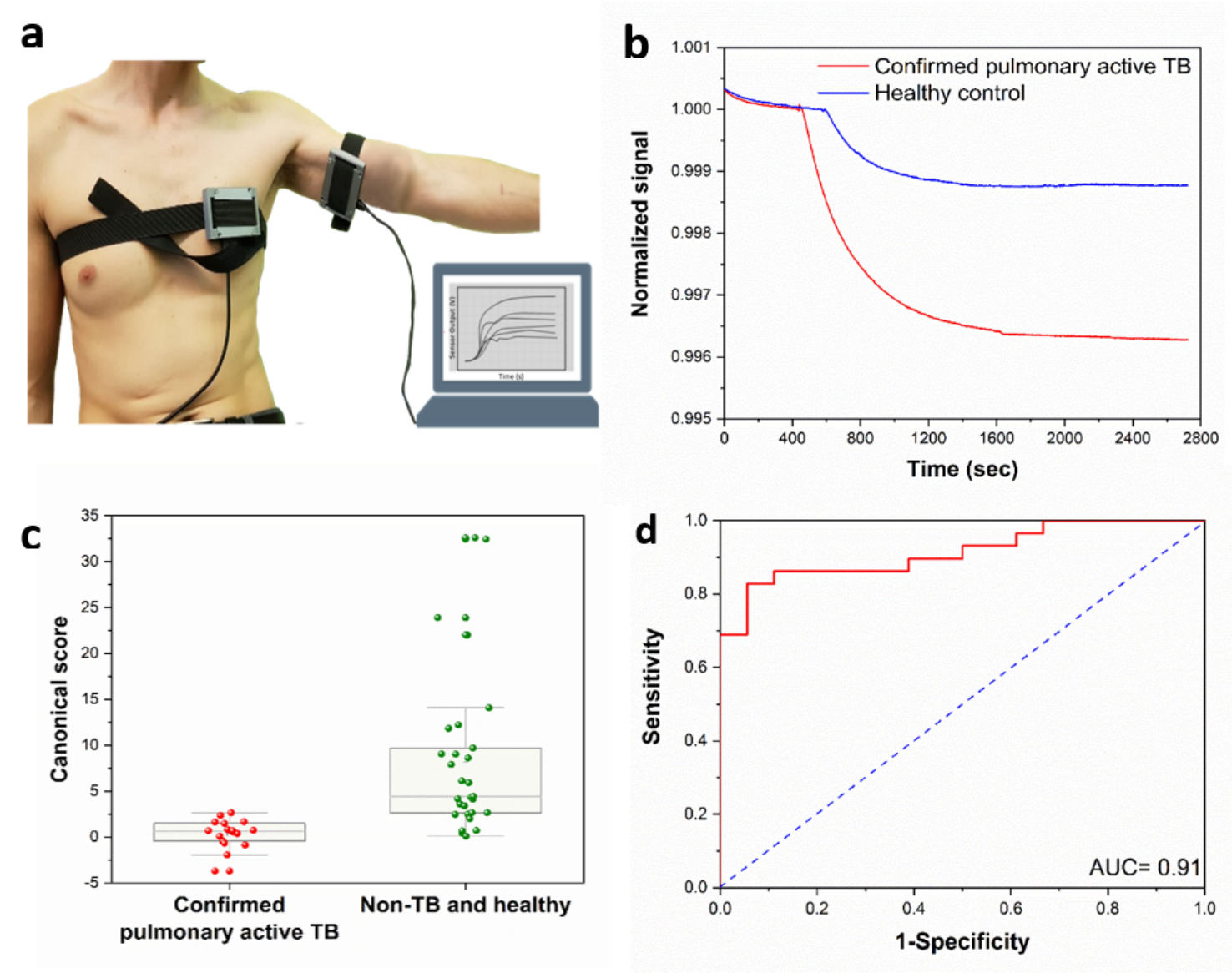

Figure 5

Wearable device TB diagnosis. a wearable devices on the chest and anterior arm of a volunteer. b representative normalized signals of one of the sensors in the wearable device attached to the anterior arm area. The plotted signal is the normalized resistance to the baseline resistance before patch is attached to the experiment participant. c boxplot of the canonical score of linear DFA model. Each point represents one sample. d ROC curve of model. AUC= area under curve.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- VishinkinSI.docx
- VishinkinSI.docx
- GCMSPeakHunterUPDATED.txt
- GCMSPeakHunterUPDATED.txt

- ReadGCMStxt22Last.txt
- ReadGCMStxt22Last.txt
- SAIndiaGCMS.xlsx
- SAIndiaGCMS.xlsx