Bridging the TB data gap: *in silico* extraction of rifampicin-resistant tuberculosis diagnostic test results from whole genome sequence data

Kamela Charmaine S. Ng[1,2], Jean Claude S. Ngabonziza[1,3], Pauline Lempens[1], Bouke Catherine de Jong[1], Frank van Leth[2,4], Conor Joseph Meehan[1,5]

[1] Mycobacteriology Unit, Department of Biomedical Sciences, Institute of Tropical Medicine, Antwerp, Belgium
[2] Amsterdam Institute for Global Health and Development, University of Amsterdam, Amsterdam, The Netherlands
[3] Department of Biomedical Services, Rwanda Biomedical Center, Kigali, Rwanda
[4] Department of Global Health, Amsterdam UMC, Location Academic Medical Center, University of Amsterdam, Amsterdam, The Netherlands
[5] School of Chemistry and Biosciences, University of Bradford, Bradford, United Kingdom

Corresponding Author:
Kamela Ng
Nationalestraat 155, Antwerp, 2000, Belgium
Email address: kng@itg.be

Abstract

Background. *Mycobacterium tuberculosis* rapid diagnostic tests (RDTs) are widely employed in routine laboratories and national surveys for detection of rifampicin-resistant (RR)-TB. However, as next generation sequencing technologies have become more commonplace in research and surveillance programs, RDTs are being increasingly complemented by whole genome sequencing (WGS). While comparison between RDTs is difficult, all RDT results can be derived from WGS data. This can facilitate longitudinal analysis of RR-TB burden regardless of the data generation technology employed. By converting WGS to RDT results, we enable comparison of data with different formats and sources particularly for countries that employ different diagnostic algorithms for drug resistance surveys. This allows national TB control programs (NTPs) and epidemiologists to exploit all available data for improved RR-TB surveillance.

Methods. We developed the Python-based TB Genome to Test (TBGT) tool that transforms WGS-derived data into laboratory-validated results of the primary RDTs – Xpert MTB/RIF, XpertMTB/RIF Ultra, Genotype MDRTB*plus* v2.0, and Genoscholar NTM + MDRTB II. The tool was validated through RDT results of RR-TB strains with diverse resistance patterns and geographic origins and applied on routine-derived WGS data.

Results. The TBGT tool correctly transformed the SNP data into the RDT results and generated tabulated frequencies of the RDT probes as well as rifampicin susceptible cases. The tool supplemented the RDT probe reactions output with the RR-conferring mutation based on identified SNPs.

Conclusion. Overall, the TBGT tool allows the NTP to assess whether currently implemented RDTs adequately detect RR-TB in their setting. With its feature to transform WGS to RDT results and enable longitudinal RR-TB data analysis, the TBGT tool may bridge the gap between and among data from periodic surveys, continuous surveillance, research, and routine tests, and may be integrated within the existing national connectivity platform for use by the NTP and epidemiologists to improve setting-specific RR-TB control. The TBGT source code and accompanying documentation is available at https://github.com/KamelaNg/TBGT.

Introduction

Resistance to rifampicin (RIF), the most potent anti-tuberculosis (TB) drug, hampers TB control. Rifampicin-resistant (RR)-TB persists as an urgent public health crisis as only 29% of estimated RR-TB patients world-wide were detected and notified in 2017. Further, 18% of previously treated TB patients were found to have RR-TB (WorldHealthOrganization 2018). The WHO endorsed the following rapid molecular RR-TB diagnostic test (RDTs) to address this concern: Xpert MTB/RIF (hereinafter called Xpert Classic) and the new version Xpert MTB/RIF Ultra (hereinafter called Ultra) [Cepheid, Sunnyvale, USA] that employ real-time polymerase chain reaction and molecular beacon technology (Blakemore et al. 2010); and the line probe assays - Genotype MDRTB*plus* v2.0 (hereinafter called LPA-Hain) [Hain Lifescience GmbH, Nehren, Germany]; and Genoscholar NTM + MDRTB II (hereinafter called LPA-Nipro) [NIPRO

Corporation, Osaka, Japan] (Dheda et al. 2017). These tests detect RR-conferring mutations within the rifampicin resistance determining region (RRDR) or hotspot of the *rpoB* gene (Andre et al. 2017; Blakemore et al. 2010; Ng et al. 2018a; Ng et al. 2018c), and revolutionized detection of RR-TB, allowing for prompt identification of patients who need to undergo adapted treatment. Xpert Classic is the most widely deployed RDT globally, implemented as the initial diagnostic tool for all presumptive pulmonary TB patients by 32 out of 48 countries in at least one of the high burden country lists (WorldHealthOrganization 2018). The wide utilization of the tests in both low and high burden TB countries resulted in the production of large volumes of RDT data, although comparisons within and between countries can be difficult due to use of differing technologies.

The utility of next generation sequencing technologies has been widely studied for improved detection of drug-resistant TB in diverse laboratory settings worldwide (Gardy & Loman 2018). Next generation sequencing-based whole genome sequencing (WGS) of *Mycobacterium tuberculosis* (*Mtb*) has been shown to accurately detect RR-TB by calling relevant single nucleotide polymorphisms (SNPs) in the *rpoB* gene (Coll et al. 2018; Miotto et al. 2017). WGS is already being widely implemented in the United Kingdom, the Netherlands, and New York, aimed towards completely replacing phenotypic drug susceptibility testing in the clinic (CRyPTICConsortium et al. 2018; de Viedma 2019). WGS implemented in high burden TB countries was shown to accurately estimate the prevalence of DR-TB (Zignol et al. 2018). The conventional periodic TB drug resistance surveys have been gathering data representative of the *Mtb* population in poor resource settings, while high income countries typically apply continuous surveillance (WorldHealthOrganization 2018; Zignol et al. 2018), to help improve the choice of standard TB treatment before full drug sensitivity profile is known (WorldHealthOrganization 2018).

The increasing use of WGS in research and public health initiatives can lead to a disconnect from the RDT-based data being generated routinely in the clinic, and the majority of surveys, widening the *Mtb* data gap. We aim to bridge this *Mtb* data gap by transforming WGS data into each of the related RDT data outputs, allowing longitudinal analysis of RR-TB prevalence and underlying mutations regardless of the switch between data generation technologies. This will allow end-users to compare data of varying formats from different technologies, and analyze previous historical strains with current isolates representative of the entire TB patient population in the country. Information obtained from this analysis may unearth years' worth of data revealing RR-TB burden, circulating RR-conferring mutations, and associated RDT probe patterns in a specific setting, critical for improved RR-TB control.

We present the TB Genome to Test (TBGT) tool, a Python 3-based set of scripts that rapidly transform WGS data type to RDT and mutation reports. The modules automatically generate frequencies of rifampicin-susceptible (RS) and RR-TB samples detected and supplement the

RDT probes output with the detected RR-conferring mutation in the format of 'wild-type amino acid-codon number-mutant amino acid'.

Materials & Methods

We developed the TBGT tool, a Python 3-executable that converts *Mtb* WGS data in the form of variant call format (VCF) or MTBseq (Kohl et al. 2018) tab files into the most likely output that would be observed from Xpert Classic, Xpert Ultra, LPA-Hain, and LPA-Nipro based on previously validated work (Ng et al. 2018a; Ng et al. 2018c). The TBGT tool can be accessed through https://github.com/KamelaNg/TBGT.

The TBGT tool

The TBGT tool can be run on any python-enabled operating system with no additional prerequisites. The TBGT workflow is shown in Figure 1. The input is a folder of files derived from a SNP calling pipeline, either in standard VCF format or tab format as output from MTBseq. The TBGT tool assumes the standard H37Rv NC000962.3 genome was used for calling these SNPs. If not, the user may remap the genome positions to the specific RR-TB-related codons using a tab delimited mapping file. An example of this tab-separated file is bundled with the tool. By default, the module will run all the RDTs on the input files.
The generated output file includes the Sample name, RIF resistance or susceptibility, the associated mutant codon position and mutation pattern, and a series of 0's or 1's indicating absence or presence of the capturing probe for the RDTs, and the RR-conferring mutation (An example output is given in Table 2). A summary table with counts and proportions of detected RS-TB cases, RDT probes, and RR-conferring mutations is also generated (Supplemental File Table S1).

Validation and sample application of the TBGT modules

We simulated VCF files for internal validation of the TBGT modules A randomly chosen VCF file was edited in Notepad ++ 7.6.3 (https://notepad-plus-plus.org/) to contain all previously tested and validated RR-conferring mutations (Andre et al. 2017; Miotto et al. 2017) mapped to known RDT results (Ng et al. 2018a; Ng et al. 2018c) (Table 1). We generated files with single and multiple RR mutations to ensure the tool is robust for all scenarios. These simulated VCFs are provided with the tool.

To show real world applicability, the TBGT tool was run on WGS of RS and RR -TB WHO Tropical Disease Research (TDR) strains with diverse resistance patterns and geographic origins stored in the Belgian Coordinated Collections of Microorganisms in the Institute of Tropical Medicine including the 47 TDR-TB strains tested in the previous validation of the RDTs against the available *rpoB* Sanger sequences of the strains (Ng et al. 2018a; Ng et al. 2018c; Vincent et al. 2012). The fastQ files (ENA accession PRJEB31023) for these samples were run through the MTBseq pipeline with default settings (Kohl et al. 2018) to generate the tab files for input to

TBGT. Additionally, WGS from 324 phenotypically RR-TB isolates of retreatment TB patients in Kinshasa, the Democratic Republic of Congo (DRC) collected in 2005-2010 (Meehan et al. 2018), and 233 WGS isolated 1991- 2010 from Rwanda were subjected to the MTBseq pipeline and run through the TBGT tool.

We were then able to extract the associated Xpert results from the 1991 to 2010 Rwandan dataset and compare it with actual Xpert results from 2012 to 2017 (Ng et al. 2018b).

Results

We tested the TBGT tool on a 64-bit Windows 10 Enterprise computer with a 2.50 GHz processor and a 8.00 GB of RAM. The running time was 39 milliseconds for 1 MTBseq tab file and 530 milliseconds for a VCF file.

The TBGT tool correctly transformed the WGS-derived SNPs in the VCF and MTBseq tab files into the laboratory-validated RDT probe reactions (Table 2), and accurately detected all all previously validated RR-conferring mutations in the simulated VCF and MTBseq clinical WGS data from the TDR, DRC (Figures 2A and 2B), and Rwanda (Figures 3A and 3B) strains, including double nucleotide changes – two SNPs covering two loci or genome positions – such as the CAC → AGC and CAC → TCC His445Ser mutations and combinations of any two unlinked mutations. The generated frequency and proportion tables (Supplemental File Table S1) showed the distribution of the RS and RR-TB strains, the RDT probe reactions, and the RR-conferring mutations detected.

The *in silico* extracted Xpert results from the Rwandan WGS collected in 1991-2010 revealed that the majority of RR-TB cases detected were linked with Xpert Classic probe E (Figure 3A). This observation was consistent with the actual Xpert results gathered in 2012 to 2017. Xpert Classic probe E was also predominantly seen in samples from Kinshasa, DRC (Figure 2A, Supplemental File Table S1). Mutation S450L, linked with Xpert Classic probe E, was the major RR-conferring mutation observed in both settings (Figures 2B and 3B).

Discussion

In this work, we developed and validated TBGT, a Python-implemented tool that extracts RR-TB RDT results from WGS-derived SNP data. If a country performed prior drug resistance surveys with a different diagnostic algorithm, the TBGT tool allows for comparing data of different formats and sources, making it possible for longitudinal analysis of previous historical data and current test results, such as periodic DRS results conducted as cross-sectional surveys, often with different technology from the previous one.

The TBGT tool allowed us to analyze almost 3 decades' worth of Xpert data from Rwanda. The longitudinal analysis of extracted Xpert results from the 1991-2010 Rwandan dataset and actual

198 Xpert results from 2012-2017 is a proof-of-concept that the TBGT tool provides an exceptional
199 opportunity to compare and analyze old and new data produced by different technologies. The
200 predominant Xpert Classic probe E observed in Rwanda is supported by the associated RR
201 mutation Ser450Leu (Tables 1 and 2) being present in the primary circulating multidrug-resistant
202 TB clone in this setting (Ngabonziza et al. 2018), and also most frequently associated with global
203 RR-TB (Cohen et al. 2015; Coll et al. 2018; Georghiou et al. 2016).
204
205 The genome-based approach of the TBGT tool also allows for reporting of disputed mutations
206 that confer occult RR and are frequently missed by the Mycobacterium Growth Indicator Tube
207 phenotypic DST (Ng et al. 2018a; Van Deun et al. 2015). For instance, disputed mutation
208 Leu452Pro, epidemiologically linked with an extensively drug-resistant TB outbreak in
209 KwaZulu-Natal, South Africa in 2005 (Cohen et al. 2015; Ioerger et al. 2009), was detected in
210 some strains from Kinshasa, DRC and Rwanda (Figures 2B and 3B, Supplemental File).
211 Mutation Leu452Pro is captured by delayed Xpert Classic probe E, denoting partially inhibited
212 probe E fluorescence (Lawn & Nicol 2011; Ng et al. 2018a), and specifically identified by the
213 unique combination of Xpert Ultra probe rpoB4B and corresponding melting temperature shift
214 (Table 1), provided sufficient *Mtb* DNA is detected. Leu452Pro was reported to be missed in
215 clinical samples by LPA-Hain due its end-probe location (Al-Mutairi et al. 2011; Rigouts et al.
216 2013), thus contributing to the RR-TB detection gap. The ability of the TBGT tool to rapidly and
217 accurately detect disputed mutations is therefore important.
218
219 As a supplementary feature, the TBGT tool picks up any RR-conferring mutation present in the
220 VCF or MTBseq tab file, and may help assess whether RDTs sufficiently detect RR-TB cases in
221 specific settings.
222
223 The TBGT tool may combine clinical data stored in large aggregated files and collected through
224 the routine laboratory and periodic survey RDT results, with WGS-derived data from national
225 surveillance programs to create a larger picture of the RS-TB and RR-TB burden in a country.
226 Through longitudinal analysis of previous historical and current RR-TB data, the TBGT tool may
227 then allow the NTPs and epidemiologists to identify unusual frequencies of RR-TB cases, RDT
228 probe reactions, and RR-conferring mutations that define a potential transmission hotspot. The
229 setting-specific RS and RR-TB counts and proportions generated by the TBGT tool may bring
230 about an extensive global database with years' worth of data for continuous statistical modeling
231 analyses and surveillance investigations.
232
233 Potentially, the TBGT tool may be integrated in the national TB diagnostic algorithm through the
234 existing connectivity platform or the newly implemented WHO cloud-based software (Dean
235 2019). The prospective application of the TBGT tool may bridge and transform the *Mtb* data gap
236 into action points for RR-TB clinicians to provide appropriate care for the individual TB/RR-TB
237 patient, and the NTP, public health officials, and policy makers to intervene at the population-

level for improved and sustained RR-TB control (Gardy & Loman 2018). The TBGT modules could also be expanded to report non-RRDR RR-causing mutations and include other TB drugs – e.g. isoniazid, pyrazinamide, and fluoroquinolones.

Conclusions

Our TBGT tool leverages on the improved access to next generation sequencing technologies in this genomic epidemiology era of TB, and complement *Mtb* WGS by rapidly transforming WGS files that store genomic sequence variations to validated outputs of the RR-TB RDTs. The prospective application of the TBGT tool within a nationwide genomic epidemiology program will bridge the *Mtb* data gap among routine RDT data, research setting WGS, and periodic survey and continuous surveillance *rpoB* sequencing and WGS data. This may result in efficient longitudinal analysis of the circulating RDT probes and underlying distribution of RR-conferring mutations. The TBGT tool may also help assess whether currently implemented RDT(s) adequately detect the RR-TB cases in the country. Linking the TBGT tool to the RR-TB algorithm we previously developed (Ng et al. 2018b) will help identify unusual frequencies of RR-TB cases as well as rare RDT probes and RR mutations that define a potential transmission hotspot, critical for active surveillance of transmitted strains in a population by the NTP.

Acknowledgements

We thank Button Ricarte for his insightful feedback on the manuscript.

ADDITIONAL INFORMATION AND DECLARATIONS

Funding

This work was supported by the Erasmus Mundus Joint Doctorate Fellowship grant 2016-1346 to K.C.S.N.

Competing Interests

The authors declare that they have no competing interests.

Author Contributions

K. C. S. Ng conceived and designed the study, wrote, implemented, troubleshooted, and finalized the scripts, generated simulated VCF files, performed internal validation and real-world application of the TBGT tool, analyzed and interpreted the data, wrote first draft of the manuscript, revised and approved final paper.

J. C. S. Ngabonziza generated MTBseq tab files of Rwanda WGS for validation of the modules, reviewed drafts of the manuscript, and approved final paper.

P. L. curated the TDR WGS, reviewed drafts of the manuscript, and approved final paper.

B. C. de Jong reviewed drafts of the manuscript and approved final paper.

F. van Leth reviewed drafts of the manuscript and approved final paper.

C. J. Meehan conceived and designed the study, refined, troubleshooted, and finalized the scripts, generated MTBseq tab files of TDR and Kinshasa, DRC WGS for validation of the modules, authored and reviewed drafts of the manuscript, and approved final paper.

Data Availability

All scripts and data for validation may be accessed through: https://github.com/KamelaNg/TBGT.

Supplemental Information

Supplemental information for this article can be found online at:

References

Al-Mutairi NM, Ahmad S, and Mokaddas E. 2011. Performance comparison of four methods for detecting multidrug-resistant Mycobacterium tuberculosis strains. Int J Tuberc Lung Dis 15:110-115.

Andre E, Goeminne L, Cabibbe A, Beckert P, Kabamba Mukadi B, Mathys V, Gagneux S, Niemann S, Van Ingen J, and Cambau E. 2017. Consensus numbering system for the rifampicin resistance-associated *rpoB* gene mutations in pathogenic mycobacteria. Clin Microbiol Infect 23:167-172. 10.1016/j.cmi.2016.09.006

Blakemore R, Story E, Helb D, Kop J, Banada P, Owens MR, Chakravorty S, Jones M, and Alland D. 2010. Evaluation of the analytical performance of the Xpert MTB/RIF assay. J Clin Microbiol 48:2495-2501. 10.1128/JCM.00128-10

Cohen KA, Abeel T, Manson McGuire A, Desjardins CA, Munsamy V, Shea TP, Walker BJ, Bantubani N, Almeida DV, Alvarado L, Chapman SB, Mvelase NR, Duffy EY, Fitzgerald MG, Govender P, Gujja S, Hamilton S, Howarth C, Larimer JD, Maharaj K, Pearson MD, Priest ME, Zeng Q, Padayatchi N, Grosset J, Young SK, Wortman J, Mlisana KP, O'Donnell MR, Birren BW, Bishai WR, Pym AS, and Earl AM. 2015. Evolution of Extensively Drug-Resistant Tuberculosis over Four Decades: Whole Genome Sequencing and Dating Analysis of Mycobacterium tuberculosis Isolates from KwaZulu-Natal. PLoS Med 12:e1001880. 10.1371/journal.pmed.1001880

Coll F, Phelan J, Hill-Cawthorne GA, Nair MB, Mallard K, Ali S, Abdallah AM, Alghamdi S, Alsomali M, Ahmed AO, Portelli S, Oppong Y, Alves A, Bessa TB, Campino S, Caws M, Chatterjee A, Crampin AC, Dheda K, Furnham N, Glynn JR, Grandjean L, Minh Ha D, Hasan R, Hasan Z, Hibberd ML, Joloba M, Jones-Lopez EC, Matsumoto T, Miranda A, Moore DJ, Mocillo N, Panaiotov S, Parkhill J, Penha C, Perdigao J, Portugal I, Rchiad Z, Robledo J, Sheen P, Shesha NT, Sirgel FA, Sola C, Oliveira Sousa E, Streicher EM, Helden PV, Viveiros M, Warren RM, McNerney R, Pain A, and Clark TG. 2018. Genome-wide analysis of multi- and extensively drug-resistant Mycobacterium tuberculosis. Nat Genet 50:307-316. 10.1038/s41588-017-0029-0

316 CRyPTICConsortium, the GP, Allix-Beguec C, Arandjelovic I, Bi L, Beckert P, Bonnet M,
317 Bradley P, Cabibbe AM, Cancino-Munoz I, Caulfield MJ, Chaiprasert A, Cirillo DM, Clifton
318 DA, Comas I, Crook DW, De Filippo MR, de Neeling H, Diel R, Drobniewski FA, Faksri K,
319 Farhat MR, Fleming J, Fowler P, Fowler TA, Gao Q, Gardy J, Gascoyne-Binzi D, Gibertoni-
320 Cruz AL, Gil-Brusola A, Golubchik T, Gonzalo X, Grandjean L, He G, Guthrie JL, Hoosdally S,
321 Hunt M, Iqbal Z, Ismail N, Johnston J, Khanzada FM, Khor CC, Kohl TA, Kong C, Lipworth S,
322 Liu Q, Maphalala G, Martinez E, Mathys V, Merker M, Miotto P, Mistry N, Moore DAJ, Murray
323 M, Niemann S, Omar SV, Ong RT, Peto TEA, Posey JE, Prammananan T, Pym A, Rodrigues C,
324 Rodrigues M, Rodwell T, Rossolini GM, Sanchez Padilla E, Schito M, Shen X, Shendure J,
325 Sintchenko V, Sloutsky A, Smith EG, Snyder M, Soetaert K, Starks AM, Supply P, Suriyapol P,
326 Tahseen S, Tang P, Teo YY, Thuong TNT, Thwaites G, Tortoli E, van Soolingen D, Walker AS,
327 Walker TM, Wilcox M, Wilson DJ, Wyllie D, Yang Y, Zhang H, Zhao Y, and Zhu B. 2018.
328 Prediction of Susceptibility to First-Line Tuberculosis Drugs by DNA Sequencing. N Engl J Med
329 379:1403-1415. 10.1056/NEJMoa1800474
330 de Viedma DG. 2019. Pathways and strategies followed in the genomic epidemiology of
331 Mycobacterium tuberculosis. Infect Genet Evol. 10.1016/j.meegid.2019.01.027
332 Dean A. 2019. Sequencing for the surveillance of drug-resistant TB. WHO Global TB
333 Programme.
334 Dheda K, Gumbo T, Maartens G, Dooley KE, McNerney R, Murray M, Furin J, Nardell EA,
335 London L, Lessem E, Theron G, van Helden P, Niemann S, Merker M, Dowdy D, Van Rie A,
336 Siu GK, Pasipanodya JG, Rodrigues C, Clark TG, Sirgel FA, Esmail A, Lin HH, Atre SR, Schaaf
337 HS, Chang KC, Lange C, Nahid P, Udwadia ZF, Horsburgh CR, Jr., Churchyard GJ, Menzies D,
338 Hesseling AC, Nuermberger E, McIlleron H, Fennelly KP, Goemaere E, Jaramillo E, Low M,
339 Jara CM, Padayatchi N, and Warren RM. 2017. The epidemiology, pathogenesis, transmission,
340 diagnosis, and management of multidrug-resistant, extensively drug-resistant, and incurable
341 tuberculosis. Lancet Respir Med. 10.1016/S2213-2600(17)30079-6
342 Gardy JL, and Loman NJ. 2018. Towards a genomics-informed, real-time, global pathogen
343 surveillance system. Nat Rev Genet 19:9-20. 10.1038/nrg.2017.88
344 Georghiou SB, Seifert M, Catanzaro D, Garfein RS, Valafar F, Crudu V, Rodrigues C, Victor
345 TC, Catanzaro A, and Rodwell TC. 2016. Frequency and Distribution of Tuberculosis
346 Resistance-Associated Mutations between Mumbai, Moldova, and Eastern Cape. Antimicrob
347 Agents Chemother 60:3994-4004. 10.1128/AAC.00222-16
348 Ioerger TR, Koo S, No EG, Chen X, Larsen MH, Jacobs WR, Jr., Pillay M, Sturm AW, and
349 Sacchettini JC. 2009. Genome analysis of multi- and extensively-drug-resistant tuberculosis from
350 KwaZulu-Natal, South Africa. PLoS One 4:e7778. 10.1371/journal.pone.0007778
351 Kohl TA, Utpatel C, Schleusener V, De Filippo MR, Beckert P, Cirillo DM, and Niemann S.
352 2018. MTBseq: a comprehensive pipeline for whole genome sequence analysis of
353 Mycobacterium tuberculosis complex isolates. PeerJ 6:e5895. 10.7717/peerj.5895

Lawn SD, and Nicol MP. 2011. Xpert(R) MTB/RIF assay: development, evaluation and implementation of a new rapid molecular diagnostic for tuberculosis and rifampicin resistance. Future Microbiol 6:1067-1082. 10.2217/fmb.11.84

Miotto P, Tessema B, Tagliani E, Chindelevitch L, Starks AM, Emerson C, Hanna D, Kim PS, Liwski R, Zignol M, Gilpin C, Niemann S, Denkinger CM, Fleming J, Warren RM, Crook D, Posey J, Gagneux S, Hoffner S, Rodrigues C, Comas I, Engelthaler DM, Murray M, Alland D, Rigouts L, Lange C, Dheda K, Hasan R, Ranganathan UDK, McNerney R, Ezewudo M, Cirillo DM, Schito M, Koser CU, and Rodwell TC. 2017. A standardised method for interpreting the association between mutations and phenotypic drug resistance in Mycobacterium tuberculosis. Eur Respir J 50. 10.1183/13993003.01354-2017

Ng KC, Meehan CJ, Torrea G, Goeminne L, Diels M, Rigouts L, de Jong BC, and Andre E. 2018a. Potential Application of Digitally Linked Tuberculosis Diagnostics for Real-Time Surveillance of Drug-Resistant Tuberculosis Transmission: Validation and Analysis of Test Results. JMIR Med Inform 6:e12. 10.2196/medinform.9309

Ng KCS, Ngabonziza JCS, Meehan CJ, Migambi P, de Jong BC, Cobelens F, and van Leth F. 2018b. Automated algorithm for early identification of rifampicin-resistant tuberculosis transmission hotspots in Rwanda. The International Journal of Tuberculosis and Lung Disease 22:S1 – S652.

Ng KCS, van Deun A, Meehan CJ, Torrea G, Driesen M, Gabriels S, Rigouts L, Andre E, and de Jong BC. 2018c. Xpert Ultra Can Unambiguously Identify Specific Rifampin Resistance-Conferring Mutations. J Clin Microbiol 56. 10.1128/JCM.00686-18

Ngabonziza JCS, Rigouts L, Mazarati JB, Uwizeye C, Nyaruhirira AU, Torrea G, de Jong BC, and Meehan CJ. 2018. Transmission drives the increase of multidrug-resistant tuberculosis in Rwanda. The International Journal of Tuberculosis and Lung Disease 22:S1 – S652.

Rigouts L, Gumusboga M, de Rijk WB, Nduwamahoro E, Uwizeye C, de Jong B, and Van Deun A. 2013. Rifampin resistance missed in automated liquid culture system for Mycobacterium tuberculosis isolates with specific *rpoB* mutations. J Clin Microbiol 51:2641-2645. 10.1128/JCM.02741-12

Van Deun A, Aung KJ, Hossain A, de Rijk P, Gumusboga M, Rigouts L, and de Jong BC. 2015. Disputed *rpoB* mutations can frequently cause important rifampicin resistance among new tuberculosis patients. Int J Tuberc Lung Dis 19:185-190. 10.5588/ijtld.14.0651

Vincent V, Rigouts L, Nduwamahoro E, Holmes B, Cunningham J, Guillerm M, Nathanson CM, Moussy F, De Jong B, Portaels F, and Ramsay A. 2012. The TDR Tuberculosis Strain Bank: a resource for basic science, tool development and diagnostic services. Int J Tuberc Lung Dis 16:24-31. 10.5588/ijtld.11.0223

WorldHealthOrganization. 2018. Global tuberculosis report.

Zignol M, Cabibbe AM, Dean AS, Glaziou P, Alikhanova N, Ama C, Andres S, Barbova A, Borbe-Reyes A, Chin DP, Cirillo DM, Colvin C, Dadu A, Dreyer A, Driesen M, Gilpin C, Hasan R, Hasan Z, Hoffner S, Hussain A, Ismail N, Kamal SMM, Khanzada FM, Kimerling M, Kohl TA, Mansjo M, Miotto P, Mukadi YD, Mvusi L, Niemann S, Omar SV, Rigouts L, Schito M,

394 Sela I, Seyfaddinova M, Skenders G, Skrahina A, Tahseen S, Wells WA, Zhurilo A, Weyer K,
395 Floyd K, and Raviglione MC. 2018. Genetic sequencing for surveillance of drug resistance in
396 tuberculosis in highly endemic countries: a multi-country population-based surveillance study.
397 Lancet Infect Dis 18:675-683. 10.1016/S1473-3099(18)30073-2.

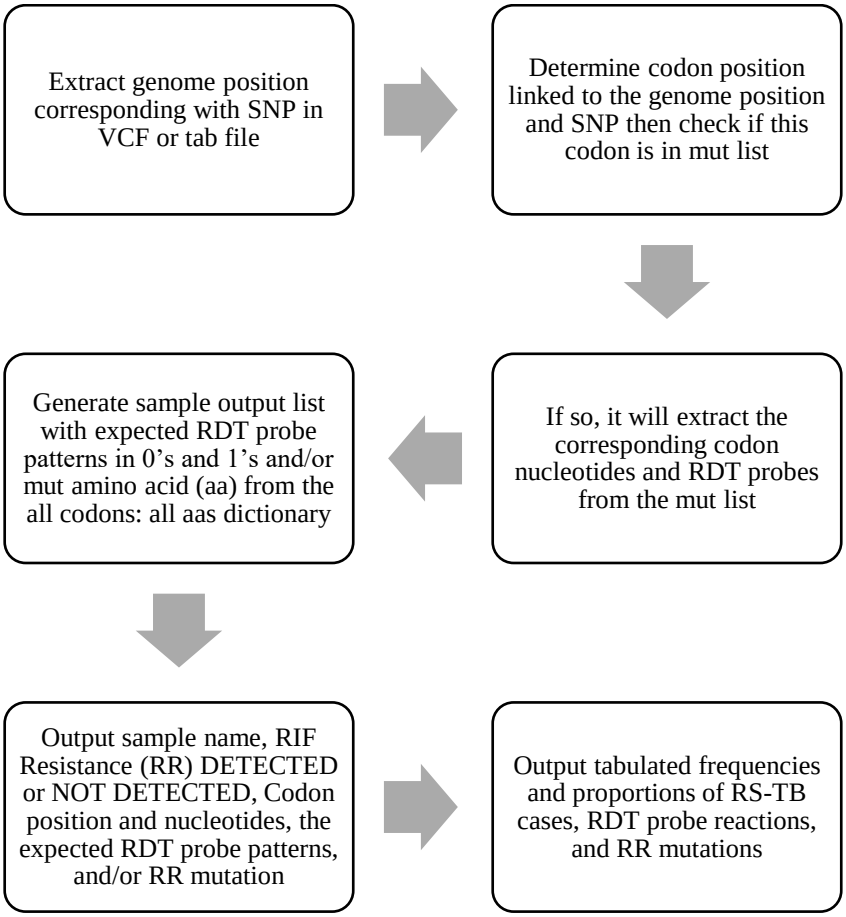

Figure 1. The TBGT tool workflow.

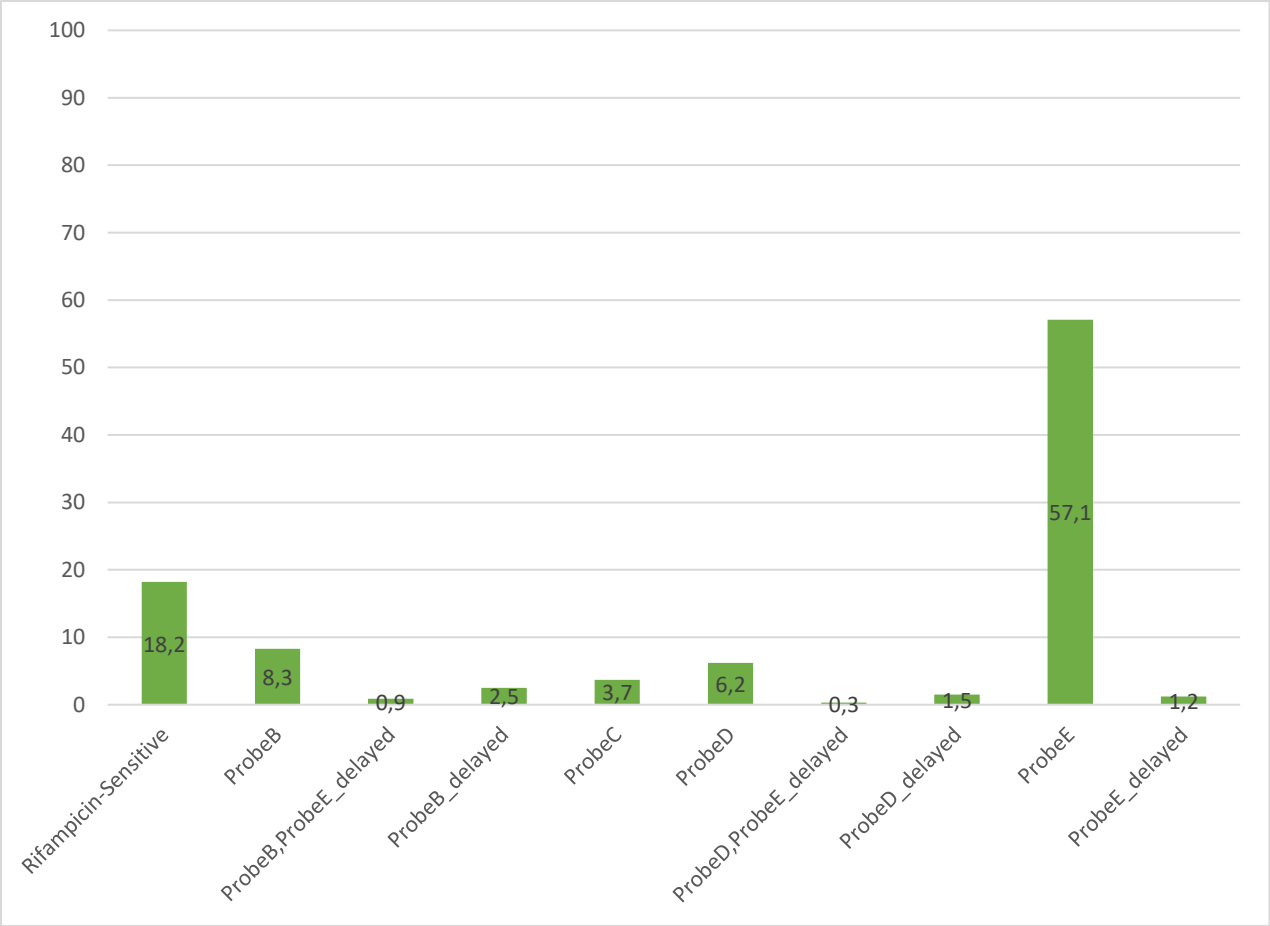

Figure 2A. Distribution of rifampicin-sensitive (RS) samples and Xpert Classic probes among rifampicin-resistant tuberculosis (RR-TB) isolates in Kinshasa, DRC from 2005 to 2010, detected by the TBGT tool.

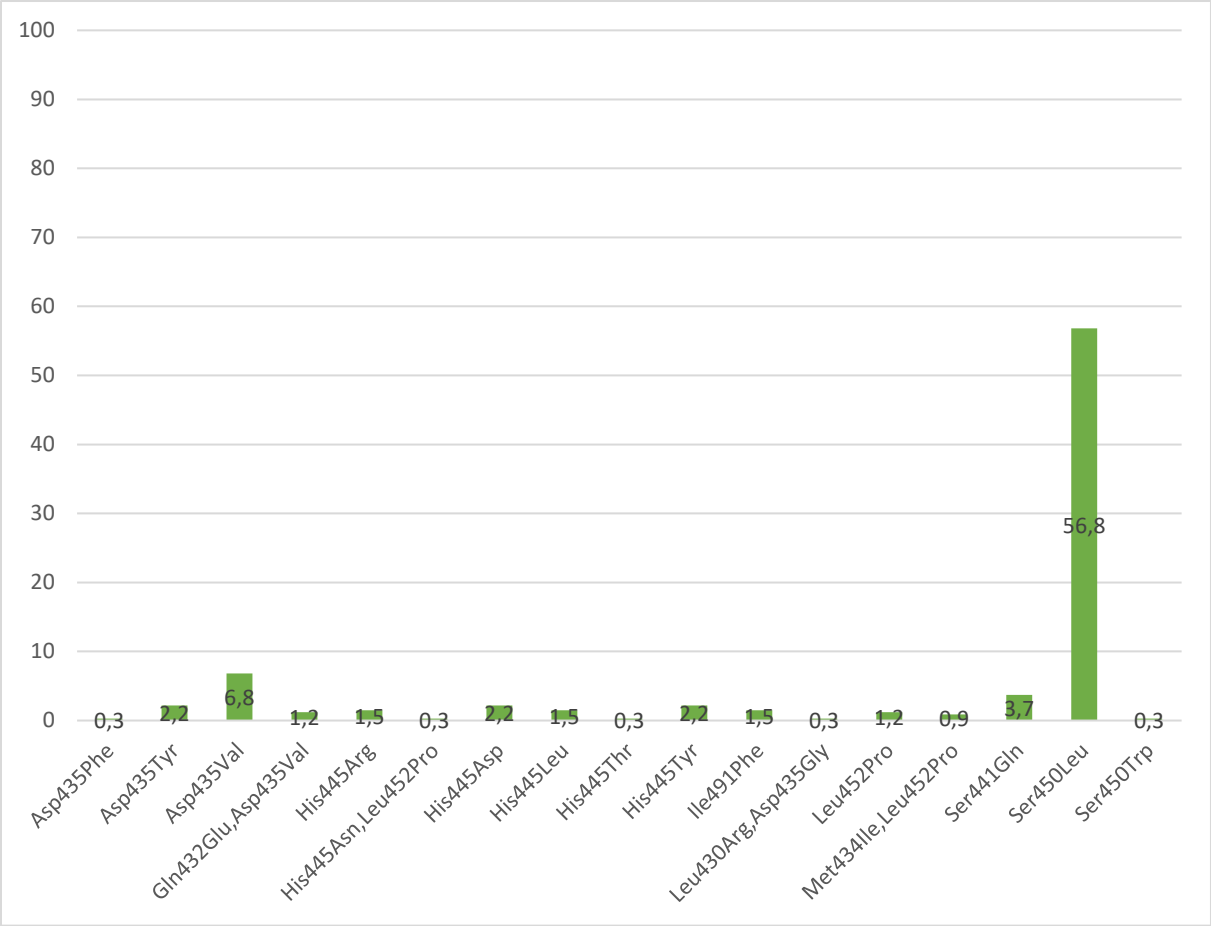

Figure 2B. Distribution of RR-conferring mutations in Kinshasa, DRC from 2005 to 2010, detected by the TBGT tool.

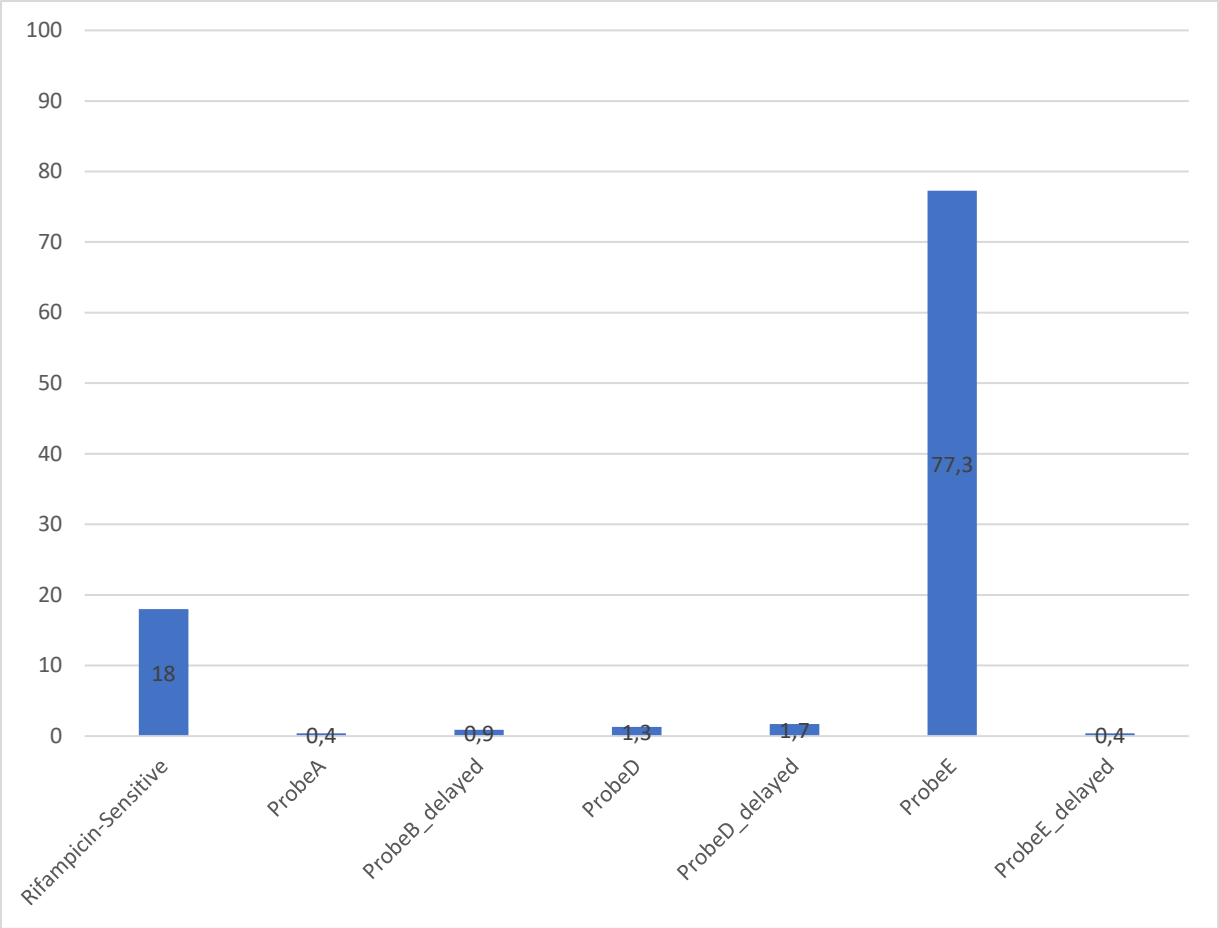

Figure 3A. Distribution of rifampicin-sensitive (RS) samples and Xpert Classic probes among rifampicin-resistant tuberculosis (RR-TB) isolates in Rwanda from 1991 to 2010, determined by the TBGT tool.

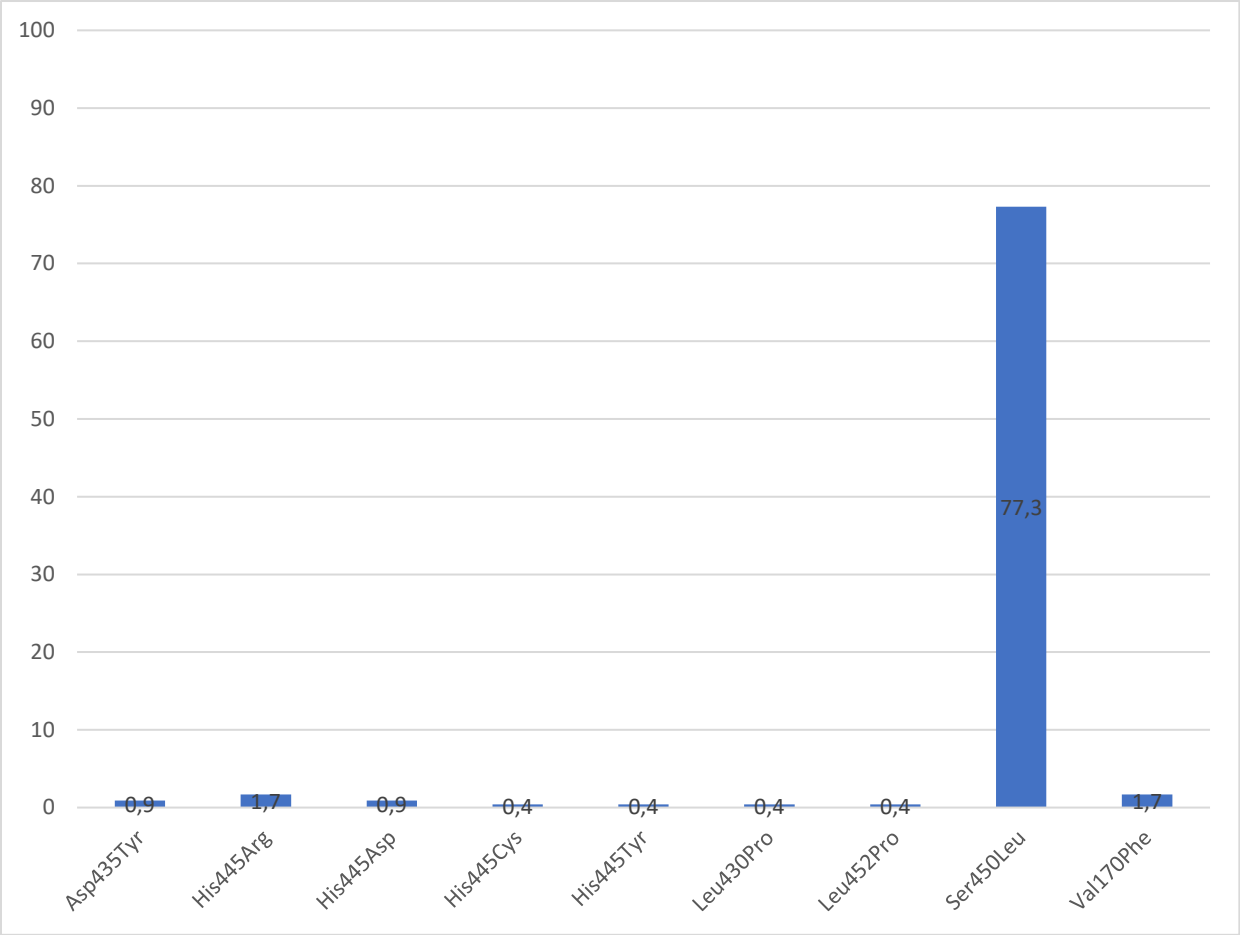

Figure 3B. Distribution of RR-conferring mutations in Rwanda from 1991 to 2010, detected by the TBGT tool.

Table 1. Rifampicin resistance (RR)-conferring mutations and associated rapid diagnostic test (RDT) results, previously tested and validated, and used as basis for developing the TBGT tool.

RR-TB RDT							RR mutation
Xpert Classic	**Xpert Ultra**		**LPA-Hain**		**LPA-Nipro**		
Capturing probe	**Capturing probe**	**Melting temperature shift**	**Absent (WT) probe**	**Developing (MUT) probe**	**Absent (WT) probe**	**Developing (MUT) probe**	
ND	ND	ND	ND	ND	ND	ND	Val170Phe
Probe A	rpoB1	3.5	WT1		S1		Ser428Arg
Probe A	rpoB1	5.9-6.3	WT2		S1		Leu430Pro
Probe A, Probe B	rpoB1	2.9	WT2		S1		Ser431Gly
Probe B	rpoB1	3.4	WT2, WT3		S1		Gln432Glu
Probe B	rpoB2	3.2	WT3		S2		Met434Ile
Probe B	rpoB2	3.3	WT3		S2		Met434Thr
Probe B	rpoB1	6.3	WT3		S2		Met434Val
Probe B	rpoB2	2.8	WT3, WT4		S2		Asp435Glu
Probe B delayed	rpoB2	5.3	WT3, WT4		S2		Asp435Phe
Probe B	rpoB2	3.3	WT3, WT4		S2		Asp435Gly
Probe B	rpoB2	3.5-3.7	WT3, WT4	MUT1	S2	R2	Asp435Val
Probe B delayed	rpoB2	4.0-4.4	WT3, WT4		S2		Asp435Tyr
Probe C	rpoB2	6.4	WT4		S2		Asn437Asp
Probe C	rpoB2; rpoB3	3.0; 2.3	WT5, WT6		S3		Ser441Leu
Probe C	rpoB2; rpoB3	4.7; 2.3	WT5, WT6		S3		Ser441Gln
Probe D	rpoB3	3.7-3.9	WT7	MUT2B	S4	R4b	His445Asp
Probe D	rpoB3	4.9	WT7		S4		His445Gly
Probe D	rpoB3	3.5-3.6	WT7		S4		His445Leu
Probe D	rpoB3	3.4-3.5	WT7		S4		His445Asn

RR-TB RDT							RR mutation
Xpert Classic	**Xpert Ultra**		**LPA-Hain**		**LPA-Nipro**		
Capturing probe	**Capturing probe**	**Melting temperature shift**	**Absent (WT) probe**	**Developing (MUT) probe**	**Absent (WT) probe**	**Developing (MUT) probe**	
Probe D	rpoB3	3.6	WT7		S4		His445Gln (CAG)
Probe D	rpoB3	4.1	WT7		S4		His445Gln (CAA)
Probe D delayed	rpoB3	1.9	WT7		S4		His445Arg
Probe D	rpoB3	4.7	WT7		S4		His445Ser
Probe D	rpoB3	4.9	WT7		S4		His445Thr
Probe D	rpoB3	3.2-3.3	WT7	MUT2A	S4	R4a	His445Tyr
Probe D	rpoB4B	5.0	WT7		S4		Lys446Gln
Probe E	rpoB3	4.0	WT8		S5		Ser450Phe
Probe E	rpoB3; rpoB4A	2.5-2.9; 6.0-6.5	WT8	MUT3	S5	R5	Ser450Leu
Probe E	rpoB3; rpoB4A	2.3-2.7; 3.3-3.7	WT8		S5		Ser450Trp
Probe E delayed	rpoB4B	5.7-6.1	WT8		S5		Leu452Pro
ND	ND	ND	ND	ND	ND	ND	Ile491Phe

Capturing probe, RDT probe associated with the RR-conferring mutation; Melting temperature shift, difference between mutant (MUT) and wild-type (WT) melting temperatures; ND, ‘not detected’, refers to mutations outside the RR determining region, not detected by the RR-TB RDTs.

Table 2. Example of combined results from the TBGT tool rapid diagnostic test (RDT) modules supplemented by the rifampicin resistance (RR)-conferring mutations detected

				Xpert Classic		Xpert Ultra		LPA-Hain		LPA-Nipro		
Filename	**RIF Resistance**	**Codon number**	**Codon**	**Capturing probe**	**Probe pattern**	**Capturing probe, Melting temperature shift**	**Probe pattern**	**Capturing probe**	**Probe pattern**	**Capturing probe**	**Probe pattern**	**RR mutation**
DRC-052577	DETECTED	450	TTG	Probe E	1 1 1 1 0 1	rpoB3,rpoB4A; 2.5-2.9,6.0-6.5	1 1 0 0 1	WT8, MUT3	1 1 1 1 1 1 1 0 0 0 0 1	S5, R5	1 1 1 1 0 0 0 0 1	Ser450Leu
DRC-091003	DETECTED	452	CCG	Probe E delayed	1 1 1 1 1 0	rpoB4B; 5.7-6.1	1 1 1 1 0	WT8	1 1 1 1 1 1 1 0 0 0 0 0	S5	1 1 1 1 0 0 0 0 0	Leu452Pro
1993-09004	DETECTED (only by *rpoB* Sanger sequencing module)	170	TTC	NOT DETECTED		NOT DETECTED		NOT DETECTED		NOT DETECTED		Val170Phe
DRC-101308	DETECTED (only by *rpoB* Sanger sequencing module)	491	TTC	NOT DETECTED		NOT DETECTED		NOT DETECTED		NOT DETECTED		Ile491Phe