JCM Accepted Manuscript Posted Online 30 May 2018
J. Clin. Microbiol. doi:10.1128/JCM.00666-18
Copyright © 2018 Doyle et al.
This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International license.

# Direct whole genome sequencing of sputum accurately identifies drug resistant *Mycobacterium tuberculosis* faster than MGIT culture sequencing

Ronan M. Doyle[1,2]*, Carrie Burgess[1], Rachel Williams[1], Rebecca Gorton[3], Helen Booth[4,5], James Brown[6,12], Josephine M. Bryant[7], Jackie Chan[8], Dean Creer[6], Jolyon Holdstock[8], Heinke Kunst[9], Stefan Lozewicz[10], Gareth Platt[3], Erika Yara Romero[1], Graham Speight[8], Simon Tiberi[9], Ibrahim Abubakar[11], Marc Lipman[6,12], Timothy D. McHugh[3], Judith Breuer[1]

1. Division of Infection and Immunity, University College London, London, UK
2. Microbiology, Virology and Infection Control, Great Ormond Street Hospital NHS Foundation Trust, London, UK
3. Centre for Clinical Microbiology, Division of Infection and Immunity, Royal Free Campus, UCL, London, UK
4. University College London Hospitals NHS Foundation Trust, London, UK
5. North Central London TB Service-South Hub, Whittington Hospital NHS Trust, London, UK
6. Royal Free London NHS Foundation Trust, London UK
7. Molecular Immunity Unit, MRC Laboratory of Molecular Biology, Department of Medicine, University of Cambridge, Cambridge, UK
8. Oxford Gene Technology, Oxford Begbroke Science Park, Begbroke, Oxfordshire, UK
9. Blizard Institute, Barts and The London School of Medicine and Dentistry, Queen Mary University of London, UK
10. North Middlesex University Hospital NHS Trust, London, UK



Accepted Manuscript Posted Online
Journal of Clinical Microbiology
JCM
Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

11. UCL Institute for Global Health, University College London, London, UK

12. UCL Respiratory, Division of Medicine, University College London, London, UK

**Running title:** WGS of drug resistant TB directly from sputum

***Corresponding Author:** Ronan M. Doyle, ronan.doyle@gosh.nhs.uk

## Abstract

The current methods available to diagnose antimicrobial resistant *Mycobacterium tuberculosis* infections require positive culture or only test a limited number of resistance-associated mutations. Rapid, accurate identification of antimicrobial resistance enables prompt initiation of effective treatment. Here, we determine the utility of whole-genome sequencing (WGS) *M. tuberculosis* directly from routinely obtained diagnostic sputum samples to provide a comprehensive resistance profile compared to Mycobacterial Growth Indicator Tube (MGIT) WGS. We sequenced *M. tuberculosis* from 43 sputum samples by targeted DNA enrichment using the Agilent SureSelectXT kit, and 43 MGIT positive samples from each participant. Thirty two (74%) sputum samples and 43 (100%) MGIT samples generated whole genomes. Time to antimicrobial resistance profile and concordance was compared with Xpert MTB/RIF and phenotypic resistance testing from culture of the same samples. Antibiotic susceptibility could be predicted from WGS of sputum within 5 days of sample receipt and up to 24 days earlier than WGS from MGIT culture and up to 31 days earlier than phenotypic testing. Direct sputum results could be reduced to 3 days with faster hybridisation and if only regions encoding drug resistance are sequenced. We show that direct sputum sequencing has the potential to provide comprehensive resistance detection



Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

significantly faster than MGIT whole genome sequencing or phenotypic testing of resistance from culture in a clinical setting. This improved turnaround time enables prompt, appropriate treatment with associated patient and health service benefit. Improvements in sample preparation are necessary to ensure comparable sensitivity and complete resistance profile predictions in all cases.

**Keywords:** *Mycobacterium tuberculosis*; Whole-genome sequencing; Pathogen DNA enrichment; Antimicrobial resistance.

## Introduction

Tuberculosis (TB) infection is a global emergency associated with an increasing burden of drug resistant *Mycobacterium tuberculosis* complex infections (1). Phenotypic testing for antimicrobial resistance detection is slow with results typically a month to six weeks after initial culture confirmation - leading to the potential for prolonged, suboptimal antibiotic treatment. Molecular assays such as the Xpert MTB/RIF (Cepheid), MTBDRplus and MTBDR*sl* (Hain Lifescience) can rapidly detect a limited number of first and second line drug resistance mutations (2). However, none are currently able to identify the full range of antibiotic resistance mutations needed for appropriately targeted therapy in people with multi-drug resistant (MDR) TB. Further, these assays recognise only a fixed number of target mutations, missing less common resistance mutations; (3) whilst Xpert MTB/RIF can only detect DNA mutations and not predict amino acid changes, resulting in potential false positives (4).



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Whole-genome sequencing (WGS) of *M. tuberculosis* allows comprehensive identification of all known drug resistant mutations for all classes of TB drugs and also can provide valuable contact tracing information (5). Recently the sequencing of organisms cultured in Mycobacterial Growth Indicator Tubes (MGIT) has been shown to be both an accurate method for detecting first and second line resistance mutations across the genome and cheaper than present routine diagnostic workflows (6). Although it is being rolled out in England, (7) it relies on bacterial culture which can delay the time to result by several weeks.

We have previously described a successful method for capturing *M. tuberculosis* DNA directly from sputum samples using biotinylated RNA baits (8). This protocol provides a possible faster alternative to sequencing *M. tuberculosis* whole genomes and could therefore offer quicker diagnosis of antibiotic resistance, leading to tailored treatment regimens with less use of antimicrobials and associated toxicity, fewer days in hospital, reduced cost and improved outcomes.

Mixed strain infections of *M. tuberculosis* are well-documented (9) and may lead to poor treatment outcomes and the possible emergence of minority drug resistant strains (10–12). Culture of *M. tuberculosis* is known to impact negatively on detection of mixtures and minority variant mutations (13), with short term MGIT culture being particularly poor at identifying mixed infections (14).

The aims of this study were: (1) to compare the utility of performing WGS directly from routinely-obtained diagnostic sputum with MGIT samples taken from the same participant (time to diagnosis plus their ability to predict antimicrobial resistance, AMR); (2) identify mixed infections and minority populations within samples.

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest



Journal of Clinical Microbiology

Accepted Manuscript Posted Online

## Materials and methods

### Study recruitment

Individuals aged 16 years or older attending a TB service with suspected pulmonary TB at seven clinics in London, UK were invited to take part in this study.

### DNA extraction

DNA was extracted from 1ml clinical samples and MGIT cultures using mechanical ribolysis and automated DNA extraction workflow. Samples were centrifuged for 30 minutes at 16,200 x g and the supernatant discarded. For MGIT cultures only, a saline prewash method was utilised to reduce the human nucleic acid component of the sample (15). 1ml of sterile saline was added to the pellet (0.9% w/v), the pellet was re-suspended and centrifuged for 15 minutes at max speed (16,200 x g). The supernatant was discarded and the process was repeated. For sputum and MGIT cultures approximately 50ul of glass beads (425-600µm) were added to each sample pellet and ribolysis was performed on a FastPrep24 platform for 45 seconds at 6.4 m/s. 240ul of extraction buffer 2 and 10ul of proteinase K was added to each sample, vortexed then incubated at 56°C for 10 minutes. DNA was extracted from samples lysates on the Diasorin IXT (Arrow) automated platform using DNA extraction cartridges eluting into 100ul.

### Quantification of extracted *Mycobacterium tuberculosis* DNA

The Xpert MTB/RIF (Cepheid) assay was performed on sputum samples as per manufacturer's instructions; reporting the *M. tuberculosis* (MTB) quantity as either very low, low, medium or high alongside $C_T$ values. The Xpert MTB/RIF assay also reported rifampicin resistance as 'detected' or 'not detected'. Drug susceptibility testing was based on phenotypic culture for first-line drugs on solid media using the resistance ratio method and was carried



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology JCM Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

out by the National Mycobacterium Reference Service using their standard protocols. A second MTB specific qPCR targeting the 16S rRNA gene (*rrs*) was utilised to quantify the MTB DNA extracted from sputum samples and MGIT cultures. For MGIT culture extracts a 1/1000 dilution was prepared prior to qPCR analysis. qPCR was performed using forward primer 5'-GTGATCTGCCCTGCACCTC-3' and reverse 5'-ATCCCACACCGCTAAAGCG-3'with a TaqMan probe ROX-AGGACCACGGGATGCATGTCTTGT-BHQ2 (16). The MTB specific qPCR reaction consisted of 12.5µl of Quantitect Muliplex NoROX mix (Qiagen), 0.2µM primers and probes and 5µl template per reaction in a total volume of 25µl. Reactions were performed in duplicate on a Rotorgene 8000 platform. PCR cycling conditions were as follows: $50^0$C for 30 mins, $95^0$C for 15 mins and 40 cycles of $94^0$C for 45 secs and $60^0$C for 45 secs. Standards were prepared from commercially sourced MTB genomic DNA (Vircell), reconstituted as directed by the manufacturers.

**Sequencing library preparation and whole genome sequencing**

Total DNA was quantified in sputum and MGIT extracts using the Qubit High Sensitivty DNA assay (Life Technologies). Carrier human genomic DNA (Promega) was added where needed to obtain a total of 200 ng of DNA input for library preparation. All DNA samples were sheared using a Covaris S2 ultrasonicator for 150 seconds (PIP 175; duty factor 5; 200 cycles per burst using frequency sweeping). Sputum samples were prepared using the SureSelectXT target enrichment system for the Illumina paired-end sequencing library protocol (Agilent Technologies). End repair, 3' addition of adenosine and ligation of adapters were all carried out according to Agilent's protocol. Prior to hybridisation, 12 cycles of precapture PCR were used using primers provided in the SureSeclectXT kit. Hybridisation of MTB DNA to the streptavidin-coated beads was carried out using a MTB specific bait set described previously (8). Briefly, 120-mer RNA baits were designed to provide non-



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

redundant coverage of the entire length of the positive strand of the H37Rv reference genome, they were synthesized by Agilent Biotechnologies. The baits can be purchased from Agilent and the bait sequences are available upon request from authors. 18 cycles of postcapture PCR were performed with indexing primers provided in the SureSelectXT kit. All Agilent recommended quality control steps were carried out. In order to compare the effect of target enrichment on MGIT sequencing, the first 14 MGIT samples were underwent library preparation using SureSelectXT and all subsequent MGIT samples had DNA libraries prepared using the NEBNext Ultra II DNA Library Prep Kit (NEB) as per the manufacturer's protocol. The resulting DNA libraries were run on either a MiSeq or NextSeq sequencer (Illumina) using either a V2 500-cycle or 500/550 Mid Output 300-cycle kit, respectively.

**Optimised sequencing method**

Three samples were prepared using SureSelectXT Fast Target Enrichment System (Agilent Technologies) as per manufacturer's protocol. A reduced bait set was designed to capture only genes associated with drug resistance and information for spoligotyping was used in hybridisation step. The reduced set of 120-mer RNA baits were synthesised by Agilent Technologies in the same way as the full set except that it only included baits that were complementary to the genes and regions in Table 1 of the H37Rv reference genome.

**Bioinformatic analysis**

Sequencing reads were trimmed for adapter content and quality using Trim Galore, keeping reads longer than 100bp. Trimmed reads were deduplicated and mapped to the H37Rv (accession: NC_000962) reference genome using BBmap allowing only successfully mapped paired reads at the 99% equivalent minimum identity across the entire read and a maximum insert size of 500bp. Duplicate mapped reads were removed using Picard tools and variants against the reference genome were called with freebayes keeping only variants with a



Journal of Clinical Microbiology

Accepted Manuscript Posted Online

minimum of 10 supporting reads, greater than 2% frequency, mapping quality greater than 20 and base quality score greater than 30, with reads present on both the forward and reverse strand, and on both the 5' or 3' end of reads. Variants found in and within 100bp of Pro-Glu (PE) and Pro-Pro-Glu (PPE) genes, mobile elements and repeat regions were discarded. For resistance calling single nucleotide variants (SNVs) were annotated using ANNOVAR (17). A maximum likelihood phylogeny was also inferred from 1113 core genome SNVs present in 64 samples representing 32 participants using RAxML (v. 8.2.1) (18) with 99 bootstrap replicates. SNV distance between pairs was calculated using R package *seqinr*. The same filtering conditions were also applied to variants for minor variant analysis. The number of reads across each variant position was normalised between pairs of samples from the same patient to adjust for the effect of read depth on variant frequency. Minor variants were filtered from the dataset if found on reads with greater sequence identity to a different MTB complex species. To further control possible contamination of paired samples with low frequency variants, MGIT and sputum samples from each pair were extracted on separate days, prepared in separate sequencing libraries and sequenced on different runs.

**Ethics approval and consent to participate**

Samples were collected with informed consent from patients attending a TB clinic setting at the participating hospitals. Approval for the study was granted by the NRES Committee East Midlands – Nottingham 1 (REC reference: 15/EM/0091). All samples were pseudo-anonymised and allocated a unique identification number.

**Data availability**

All sequence data associated with this study has been deposited in the European Nucleotide Archive under study accession number PRJEB21685.



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology JCM Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

## Results

### Genomic coverage

Sixty three participants were prospectively enrolled. A paired sputum and MGIT sample was sequenced from 43 patients. This is due to ten participants not having a MGIT sample collected for sequencing, another eight samples being smear and Xpert MTB/RIF negative, and two where the volume of sputum was insufficient for DNA extraction. Samples sequenced from MGIT culture had a higher reference genome coverage as compared with those obtained directly from sputum (Fig. 1) and this was correlated to the increased *M. tuberculosis* DNA available from the former (Fig S1). Enrichment of MGIT culture samples using the *M. tuberculosis* probes also enhanced the quality and depth of sequence (Fig S1).

We next evaluated whether bacterial load, as measured by smear and Xpert MTB/RIF, could be used to predict the success of whole genome sequencing. From 43 patients, 32 sputum samples (74.4%) and 43 MGIT samples (100%) generated whole genomes (>85% coverage against reference genome) (Fig. 2A). Sputum sequencing success was linked to estimated input pathogen copy number. We stratified participants into 16 with high (3+ smear result, Xpert MTB/RIF High), 18 with medium (2+ smear result, Xpert MTB/RIF Medium) and 9 low bacterial load (scanty or 1+ smear result, Xpert MTB/RIF Low) and found 87.5% of sputum samples with high bacterial load samples generated complete genomes as compared to 72.2% with medium and 55.5% with low bacterial load (Fig. 2B & 2C). We were also able to recover partial genomes for two sputum samples that were reported as negative by smear microscopy but Xpert MTB/RIF positive.

### MGIT and sputum sequence variation



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Comparison of the 32 patients with both complete sputum and MGIT genomes available, showed no unique consensus sequence variation between the pairs. The identity between sputum and MGIT consensus sequences is shown in a heat map (Fig. 3) and phylogenetic tree (Fig. S2). Twenty three MGIT and sputum pairs showed no SNV between them at the consensus level, while nine patients' sample pairs differed by one or two single nucleotides. In all nine patients the consensus polymorphism was present as a minority variant in the matched sputum or MGIT sample (Table S1).

**Time to antibiotic resistance prediction**

Allowing for sample batching, which was only carried out for study purposes, direct sequencing of sputum using targeted enrichment reduced the time to antibiotic susceptibility prediction initially to five days, as compared with a mean of 11 (s.d. 6) days for MGIT sequencing (Fig. 4). This was reduced further by protocol optimisation to three days when a reduced bait set was used that captures only the regions with putative resistance mutations (Table 1 & Fig. 4). Hybridisation optimisation could also reduce the whole genome protocol to 4 days (Fig. 4). The reduced bait set targeted 35,960bp of the H37Rv reference genome in total and successfully re-sequenced three MDR-TB samples (noted in Fig 4.) from this study to a high average depth of coverage (>2,000X) over the captured regions. All eight genotypic resistant variants identified in the whole genome sequencing data were also identified after re-sequencing with the reduced bait set (Fig. S3). Overall, 36 sputum samples with complete genomes, including 77% of those with drug resistance mutations, would have been reported a mean of 9 days earlier than MGIT sequencing and a mean of 35 days earlier than phenotypic testing using the optimised three day protocol.

**Antibiotic resistance concordance**

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest



Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

We found complete concordance between resistance mutations identified in paired MGIT and sputum samples from nine participants when there was >85% single read coverage against reference genome. Four participants missed resistance mutations where sputum sequencing read coverage was too low to make a reliable call (Table 2). Xpert MTB/RIF, sputum WGS, MGIT WGS were concordant with phenotypic resistance testing in 21 out of 23 resistance mutations identified (Table 2). The exceptions were where a variant in participant RF015GT predicting an amino acid change Ser428Iso in *rpoB* (H37Rv codon numbering. *Escherichia coli rpoB* numbering S509I) reported as resistant by Xpert MTB/RIF, but the reference laboratory found it to be susceptible (Table 2). Previous publications have shown not all SNVs at this position are associated with resistance (19–21). A fixed mutation in *f* predicting an amino acid change Ser315Thr was confirmed to confer high levels of resistance to isoniazid in five samples, but patient BH052SA with the same mutation was found to be susceptible. This common polymorphism in *katG* has been previously been shown to confer consistently high levels of isoniazid resistance to *M. tuberculosis* (22–24). Whole genome sequencing from both MGIT and sputum samples also identified streptomycin (three patients) and Para-aminosalicylic acid (PAS, four patients) resistance mutations, neither of which is routinely tested within the phenotypic assay in the UK.

**Mixed infections and minority variants**

No mixed infections were detected. Using data normalised for read depth, from 32 matched sputa-MGIT samples, minor frequency variation was low with only 88 minority bi-allelic sites meeting the quality criteria identified in all samples, representing 0.002% unique variable positions across the genomes. We undertook stringent procedures to exclude sequencing error, the presence of closely related *M. tuberculosis* complex species in sputum, contamination in and between sequencing runs, and lab contamination after sample collection as potential causes for the findings. None of the variant alleles were at positions known to be



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

associated with antimicrobial resistance. In 41% (13) of cases directly sequenced sputa had higher numbers of novel minority variants identified than the matched MGIT as compared to 19% (6) of MGIT samples with more minority variants in the sputum (Fig. S4).

To control for the potential influence of the SureSelectXT step, we analysed the proportion of minority variants shared between sputum and eight SureSelectXT enriched MGIT samples (40%) as compared between sputum and 16 non-enriched samples (38%) and found no statistically significant difference (p=0.854). Overall, 37.2% of minority variants were concordant between sputum and MGIT and the read frequencies with which they occurred were weakly correlated (Fig. S5). This correlation was skewed by one patient (WH044IL) whose MGIT and sputum samples were both more variable than other samples (Fig. S4) and in whom seven variants were at much higher frequency in the MGIT than in sputum (at 30-45% versus ~5% respectively, Fig, S5 & Fig. S6). However, there was no evidence of mixed genotypes. The five synonymous minor variants and two non-synonymous variants occurring in two genes of unknown function (Rv3529c and Rv3888c), were distributed across the genome, and were not shared across any other pairs of samples.

## Discussion

We have shown that using target enrichment WGS methodology directly from diagnostic sputum samples generates resistance data, at most, up to 24 days earlier than MGIT culture WGS and up to 31 days faster than phenotypic testing of *Mycobacterium tuberculosis*. Sputum sequencing only achieved whole genome sequences suitable for predicting resistance mutations in 32/43 (74%), though this included a smear negative sputum sample. Our demonstration that the quality of sequence data is strongly correlated with the input level of TB DNA (Fig, S1), means that the success of sequencing can be predicted using semi-

Accepted Manuscript Posted Online

quantitative methods such as smear microscopy and Xpert MTB/RIF (Fig. 2), notwithstanding their variable performance (25).

Our data compare well with a recent report describing WGS of sputum where contaminating human DNA had been depleted (26). While this method achieved a slightly faster turnaround time on diagnostic samples (2 versus 5 days), it may be less susceptible than targeted enrichment, as only 60% (24/40) of smear positive samples yielded sequence data suitable for resistance prediction. The study did not report bacterial load so a thorough comparison of sensitivity cannot be performed. The methods are, however, highly complementary and combining the two would likely improve genome copy input and increase direct WGS sensitivity. Our experience of enrichment methods (27, 28) also predicts that redesign of the first generation probe set would further improve detection of resistance mutations.

Direct sequencing of sputum is currently slower than rapid methods such as the Xpert MTB/RIF and the Hain MTBDRplus and MTBDRsl assays for detecting resistance. However there are major advantages using WGS. First, unlike existing rapid methods, it can accurately identify the precise nucleotide change causing resistance. In our study Xpert MTB/RIF reported resistance in a susceptible organism where there was a nucleotide change at position 428 in the *rpoB* gene not associated with resistance (19–21). Where discordant resistance results were found between molecular methods and the routine phenotypic testing, we could not repeat the phenotypic testing as this was carried out historically by a centralised reference laboratory. Therefore these discrepancies cannot be confirmed and this represents a limitation of this type of study. Second, WGS, unlike rapid methods which target specific mutations, is able to detect resistance mutations for a wider range of second and third line drugs, and also new drugs where current rapid tests would require costly redesign. This has already helped us personalise treatment in a case of drug resistant *M. tuberculosis* (29). Third, the data from



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology JCM Journal of Clinical Microbiology

Accepted Manuscript Posted Online

direct WGS of sputum can report evolutionary relationships between samples (Fig. S2) - providing the most detailed information on transmission dynamics available.

An important objective of our study was to evaluate the potential for direct sequencing from sputum to detect mixed *M. tuberculosis* infections which are sub-optimally identified by MGIT and solid culture (13, 14). Mixed infections are important in the pathology of TB and the ability to detect resistance variants that are not at consensus level, although not necessarily common (30), can affect antibiotic stewardship (31). Mixed infections and MDR-TB are more prevalent in countries with a much higher burden of TB than the UK and a greater prevalence of drug resistance, whilst high levels of HIV amplify the problem (32, 33). We were able to detect significantly more minority single nucleotide variants (SNVs) in sputum compared to the matched MGIT sequence (Fig. S5), despite the mostly clonal populations in this study and the greater read depths achieved from MGIT sequencing. The origin of this heterogeneity remains unconfirmed, although we rigorously excluded contamination and methodological error. SNVs could be due to the presence in sputum of non-tuberculous mycobacteria and other species which are known to have sequence homology with *M. tuberculosis* and may theoretically be detected by targeted enrichment. However, our use of highly stringent sequence mapping and the fact that the SNVs were detected across the genome and not concentrated in regions generally associated with cross hybridisation, suggests that they are real. In case WH044IL, seven SNVs present in sputum increased in frequency in MGIT culture, possibly reflecting a selective growth advantage for this haplotype, particularly as one non-synonymous SNV occurred in the Rv3888c gene which has been shown to be essential for mycobacterial *in vitro* growth (34). This result confirms suggestions that diversity is lost and that culture-related selection of some variants can occur even during limited MGIT culture. Thus MGIT culture may not be representative of the original sample, and could potentially reduce the likelihood of identifying low level



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

resistance mutations and mixed infections that may act as a reservoir for resistance development.

The standard diagnostic workflows for *M. tuberculosis* are costly and time-consuming. Ground breaking work within Public Health England has demonstrated that sequencing *M. tuberculosis* whole genomes from positive MGIT cultures is faster and cheaper (6) and where sputum pathogen DNA concentration is low, early MGIT sequencing could still be the best possible workaround (15). Direct sequencing of *M. tuberculosis* from sputum has the potential to reduce the time to antimicrobial resistance detection within a clinically relevant timeframe (26). We show here that its success is critically dependent on the input genome copies of pathogen DNA. While enrichment increases the cost of pathogen sequencing, this could be offset, as demonstrated in our study, by only enriching areas of interest on the genome. It is important to note that the infrastructure and expertise for rapid, high throughput, targeted enrichment sequencing directly from clinical material of *M. tuberculosis* and other pathogens already exists in genomic centres where cancer and genetic disease sequencing uses this methodology. We believe, therefore, that effective scale-up and implementation of this rapid and accurate technology is relatively easy, once the technique has been optimised.

**Acknowledgements**

We would like to thank all participants who agreed to take part of this study. We would also like to thank Lusha Kellgren, Maria McEwan, Narinder Boparai, Jacqui White, Neil Jones, Stephen Morris-Jones, Surjo De and Trupti Patel, for collection of samples at The Whittington Health NHS Trust, University College London Hospitals and the TB Service North Central London. Nirmala Ghimire and Devan Vaghela, for collection of samples at



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology
JCM
Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Bart's Health NHS Trust. Josephine Silles, Pragna Patel, Marinos Christofi, Thomas Rendal, Yemi Martins and Esmeralda Sicat for the collection of samples at North Middlesex University Hospital. This work was funded by the UCLH/UCL NIHR Biomedical Resource Centre (Grant number: BRC/176/III/JB/101350) and the PATHSEEK European Union's Seventh Programme for research, technological development and demonstration (grant number 304875). The authors declare that they have no competing interests.

**References**

1. WHO. 2017. Global tuberculosis report 2017. WHO.

2. Scott LE, McCarthy K, Gous N, Nduna M, Rie AV, Sanne I, Venter WF, Duse A, Stevens W. 2011. Comparison of Xpert MTB/RIF with Other Nucleic Acid Technologies for Diagnosing Pulmonary Tuberculosis in a High HIV Prevalence Setting: A Prospective Study. PLOS Med 8:e1001061.

3. Jenkins C, Claxton AP, Shorten RJ, McHugh TD, Gillespie SH. 2005. Rifampicin Resistance in Tuberculosis Outbreak, London, England. Emerg Infect Dis 11:931.

4. Ocheretina O, Byrt E, Mabou M-M, Royal-Mardi G, Merveille Y-M, Rouzier V, Fitzgerald DW, Pape JW. 2016. False-positive rifampin resistant results with Xpert MTB/RIF version 4 assay in clinical samples with a low bacterial load. Diagn Microbiol Infect Dis 85:53–55.

5. Witney AA, Gould KA, Arnold A, Coleman D, Delgado R, Dhillon J, Pond MJ, Pope CF, Planche TD, Stoker NG, Cosgrove CA, Butcher PD, Harrison TS, Hinds J. 2015.



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology JCM Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Clinical Application of Whole-Genome Sequencing To Inform Treatment for Multidrug-Resistant Tuberculosis Cases. J Clin Microbiol 53:1473–1483.

6. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: a prospective study. Lancet Respir Med 4:49–58.

7. Walker TM, Cruz ALG, Peto TE, Smith EG, Esmail H, Crook DW. 2017. Tuberculosis is changing. Lancet Infect Dis 17:359–361.

8. Brown AC, Bryant JM, Einer-Jensen K, Holdstock J, Houniet DT, Chan JZM, Depledge DP, Nikolayevskyy V, Broda A, Stone MJ, Christiansen MT, Williams R, McAndrew MB, Tutill H, Brown J, Melzer M, Rosmarin C, McHugh TD, Shorten RJ, Drobniewski F, Speight G, Breuer J. 2015. Rapid Whole-Genome Sequencing of Mycobacterium tuberculosis Isolates Directly from Clinical Samples. J Clin Microbiol 53:2230–2237.

9. Cohen T, van Helden PD, Wilson D, Colijn C, McLaughlin MM, Abubakar I, Warren RM. 2012. Mixed-Strain Mycobacterium tuberculosis Infections and the Implications for Tuberculosis Treatment and Control. Clin Microbiol Rev 25:708–719.

10. van Rie A, Victor TC, Richardson M, Johnson R, van der Spuy GD, Murray EJ, Beyers N, Gey van Pittius NC, van Helden PD, Warren RM. 2005. Reinfection and mixed infection cause changing Mycobacterium tuberculosis drug-resistance patterns. Am J Respir Crit Care Med 172:636–642.

11. Baldeviano-Vidalón GC, Quispe-Torres N, Bonilla-Asalde C, Gastiaburú-Rodriguez D, Pro-Cuba JE, Llanos-Zavalaga F. 2005. Multiple infection with resistant and sensitive



Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

M. tuberculosis strains during treatment of pulmonary tuberculosis patients. Int J Tuberc Lung Dis Off J Int Union Tuberc Lung Dis 9:1155–1160.

12. Zetola NM, Shin SS, Tumedi KA, Moeti K, Ncube R, Nicol M, Collman RG, Klausner JD, Modongo C. 2014. Mixed Mycobacterium tuberculosis complex infections and false-negative results for rifampin resistance by GeneXpert MTB/RIF are associated with poor clinical outcomes. J Clin Microbiol 52:2422–2429.

13. Martín A, Herranz M, Ruiz Serrano MJ, Bouza E, García de Viedma D. 2010. The clonal composition of Mycobacterium tuberculosis in clinical specimens could be modified by culture. Tuberc Edinb Scotl 90:201–207.

14. Hanekom M, Streicher EM, Berg DV de, Cox H, McDermid C, Bosman M, Pittius NCG van, Victor TC, Kidd M, Soolingen D van, Helden PD van, Warren RM. 2013. Population Structure of Mixed Mycobacterium tuberculosis Infection Is Strain Genotype and Culture Medium Dependent. PLOS ONE 8:e70178.

15. Votintseva AA, Pankhurst LJ, Anson LW, Morgan MR, Gascoyne-Binzi D, Walker TM, Quan TP, Wyllie DH, Del Ojo Elias C, Wilcox M, Walker AS, Peto TEA, Crook DW. 2015. Mycobacterial DNA extraction for whole-genome sequencing from early positive liquid (MGIT) cultures. J Clin Microbiol 53:1137–1143.

16. Honeyborne I, McHugh TD, Phillips PPJ, Bannoo S, Bateson A, Carroll N, Perrin FM, Ronacher K, Wright L, van Helden PD, Walzl G, Gillespie SH. 2011. Molecular bacterial load assay, a culture-free biomarker for rapid and accurate quantification of sputum Mycobacterium tuberculosis bacillary load during treatment. J Clin Microbiol 49:3905–3911.



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Accepted Manuscript Posted Online

17. Yang H, Wang K. 2015. Genomic variant annotation and prioritization with ANNOVAR and wANNOVAR. Nat Protoc 10:1556–1566.

18. Stamatakis A. 2014. RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies. Bioinformatics 30:1312–1313.

19. Williams DL, Waguespack C, Eisenach K, Crawford JT, Portaels F, Salfinger M, Nolan CM, Abe C, Sticht-Groh V, Gillis TP. 1994. Characterization of rifampin-resistance in pathogenic mycobacteria. Antimicrob Agents Chemother 38:2380–2386.

20. Sekiguchi J, Miyoshi-Akiyama T, Augustynowicz-Kopeć E, Zwolska Z, Kirikae F, Toyota E, Kobayashi I, Morita K, Kudo K, Kato S, Kuratsuji T, Mori T, Kirikae T. 2007. Detection of Multidrug Resistance in Mycobacterium tuberculosis. J Clin Microbiol 45:179–192.

21. Zenteno-Cuevas R, Zenteno JC, Cuellar A, Cuevas B, Sampieri CL, Riviera JE, Parissi A. 2009. Mutations in rpoB and katG genes in Mycobacterium isolates from the Southeast of Mexico. Mem Inst Oswaldo Cruz 104:468–472.

22. Wengenack NL, Uhl JR, St Amand AL, Tomlinson AJ, Benson LM, Naylor S, Kline BC, Cockerill FR, Rusnak F. 1997. Recombinant Mycobacterium tuberculosis KatG(S315T) is a competent catalase-peroxidase with reduced activity toward isoniazid. J Infect Dis 176:722–727.

23. Saint-Joanis B, Souchon H, Wilming M, Johnsson K, Alzari PM, Cole ST. 1999. Use of site-directed mutagenesis to probe the structure, function and isoniazid activation of the catalase/peroxidase, KatG, from Mycobacterium tuberculosis. Biochem J 338 ( Pt 3):753–760.



Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

444 24. Pym AS, Saint-Joanis B, Cole ST. 2002. Effect of katG mutations on the virulence of
445 Mycobacterium tuberculosis and the implication for transmission in humans. Infect
446 Immun 70:4955–4960.

447 25. Devonshire AS, O'Sullivan DM, Honeyborne I, Jones G, Karczmarczyk M, Pavšič J,
448 Gutteridge A, Milavec M, Mendoza P, Schimmel H, Van Heuverswyn F, Gorton R,
449 Cirillo DM, Borroni E, Harris K, Barnard M, Heydenrych A, Ndusilo N, Wallis CL,
450 Pillay K, Barry T, Reddington K, Richter E, Mozioğlu E, Akyürek S, Yalçınkaya B,
451 Akgoz M, Žel J, Foy CA, McHugh TD, Huggett JF. 2016. The use of digital PCR to
452 improve the application of quantitative molecular diagnostic methods for tuberculosis.
453 BMC Infect Dis 16:366.

454 26. Votintseva AA, Bradley P, Pankhurst L, Del Ojo Elias C, Loose M, Nilgiriwala K,
455 Chatterjee A, Smith EG, Sanderson N, Walker TM, Morgan MR, Wyllie DH, Walker
456 AS, Peto TEA, Crook DW, Iqbal Z. 2017. Same-day diagnostic and surveillance data
457 for tuberculosis via whole genome sequencing of direct respiratory samples. J Clin
458 Microbiol.

459 27. Depledge DP, Palser AL, Watson SJ, Lai IY-C, Gray ER, Grant P, Kanda RK, Leproust
460 E, Kellam P, Breuer J. 2011. Specific capture and whole-genome sequencing of viruses
461 from clinical samples. PloS One 6:e27805.

462 28. Christiansen MT, Brown AC, Kundu S, Tutill HJ, Williams R, Brown JR, Holdstock J,
463 Holland MJ, Stevenson S, Dave J, Tong CYW, Einer-Jensen K, Depledge DP, Breuer J.
464 2014. Whole-genome enrichment and sequencing of Chlamydia trachomatis directly
465 from clinical samples. BMC Infect Dis 14:591.



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology
JCM
Journal of Clinical Microbiology

Accepted Manuscript Posted Online

466 29. Nimmo C, Doyle R, Burgess C, Williams R, Gorton R, McHugh TD, Brown M, Morris-
467 Jones S, Booth H, Breuer J. 2017. Rapid identification of a Mycobacterium tuberculosis
468 full genetic drug resistance profile through whole genome sequencing directly from
469 sputum. Int J Infect Dis 62:44–46.

470 30. Witney AA, Bateson ALE, Jindani A, Phillips PPJ, Coleman D, Stoker NG, Butcher
471 PD, McHugh TD, RIFAQUIN Study Team. 2017. Use of whole-genome sequencing to
472 distinguish relapse from reinfection in a completed tuberculosis clinical trial. BMC Med
473 15:71.

474 31. Cohen T, Chindelevitch L, Misra R, Kempner ME, Galea J, Moodley P, Wilson D.
475 2016. Within-Host Heterogeneity of Mycobacterium tuberculosis Infection Is
476 Associated With Poor Early Treatment Response: A Prospective Cohort Study. J Infect
477 Dis 213:1796–1799.

478 32. Bifani PJ, Mathema B, Kurepina NE, Kreiswirth BN. 2002. Global dissemination of the
479 Mycobacterium tuberculosis W-Beijing family strains. Trends Microbiol 10:45–52.

480 33. Eldholm V, Rieux A, Monteserin J, Lopez JM, Palmero D, Lopez B, Ritacco V, Didelot
481 X, Balloux F. 2016. Impact of HIV co-infection on the evolution and transmission of
482 multidrug-resistant tuberculosis. eLife 5:e16644.

483 34. Griffin JE, Gawronski JD, DeJesus MA, Ioerger TR, Akerley BJ, Sassetti CM. 2011.
484 High-Resolution Phenotypic Profiling Defines Genes Essential for Mycobacterial
485 Growth and Cholesterol Catabolism. PLOS Pathog 7:e1002251.

486

487 **Tables**



Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Journal of Clinical Microbiology
JCM
Journal of Clinical Microbiology

Accepted Manuscript Posted Online

488 **Table 1. Target genes/regions and reason for inclusion in the reduced *M. tuberculosis***

489 **bait set.**

| **Gene target** | **Region property** |
|---|---|
| *gyrB/gyrA* | Fluoroquinolone resistance |
| *rpoB/rpoC* | Rifampicin resistance |
| *rpsL* | Streptomycin resistance |
| *Rrs* | Streptomycin, amikacin and kanamycin resistance |
| *gidB* | Streptomycin resistance |
| *mabA/fabG1* + promoter + *inhA* | Isoniazid and ethionamide resistance |
| *katG* | Isoniazid resistance |
| *kasA* | Isoniazid resistance |
| *aphC-oxyR* | Isoniazid resistance |
| *tlyA* | Capreomycin resistance |
| Promoter + *pncA* | Pyrazinamide resistance |
| *eis* + promoter | Kanamycin resistance |
| *thyA* | PAS resistance |
| *embC* | Ethambutol resistance |
| *embB* | Ethambutol resistance |
| *ethA* | Ethionamide resistance |
| Direct repeat locus | Spoligotyping |

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest



Journal of Clinical Microbiology
JCM
Journal of Clinical Microbiology

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

**Table 2. Antimicrobial resistance profiles from 13 patients with evidence of resistance from direct sputum sequencing using whole genome bait set.**

| | | RIFAMPICIN | | | | | | | | ISONIAZID | | | | | | | ETHAMBUTOL | | | | | PYRAZINAMIDE | | | STREPTOMYCIN | | | | | PAS | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | Smear result | rpoB:c.C1337T:p.H445Y | | rpoB:c.1333_1334TG | | rpoB:c.C1349T:p.S450L | | GeneXpert RIF resistance | Solid culture phenotype | katG:c.G944C:p.S315T | | fabG1:c.-8 | | inhA:c.T280G:p.S94A | | Solid culture phenotype | embB:c.G1217A:p.G406D | | embB:c.A1490G:p.Q497R | | Solid culture phenotype | pncA:c.T11C:p.L4S | | Solid culture phenotype | rpsL:c.A128G:p.K43R | | gidc.:102delG:p.G34fs | | Solid culture phenotype | thyA:c.A604G:p.T202A | | Solid culture phenotype |
| **Patient** | | Sputum | MGIT | Sputum | MGIT | Sputum | MGIT | | | Sputum | MGIT | Sputum | MGIT | Sputum | MGIT | | Sputum | MGIT | Sputum | MGIT | | Sputum | MGIT | | Sputum | MGIT | Sputum | MGIT | | Sputum | MGIT | |
| **BH001MC** | + | S | ND | S | ND | S | ND | ND | S | S | ND | S | ND | S | ND | S | S | ND | S | ND | S | ND | S | S | S | ND | S | ND | ND | R | ND | ND |
| **BH041OS** | + | . | S | . | S | . | R | R | R | . | R | . | S | . | S | R | . | S | . | R | R | . | R | R | . | R | . | S | ND | . | S | ND |
| **BH052SA** | +++ | S | S | S | S | S | S | S | S | R | R | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | ND | S | S | ND |
| **BH056ESDS** | + | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | S | ND | R | R | ND |
| **RF002AH** | + | S | S | S | S | S | S | ND | S | R | R | S | S | S | S | R | S | S | S | S | S | S | S | S | S | S | S | S | ND | S | S | ND |
| **RF009ZC** | Neg | . | S | . | S | . | S | S | S | . | R | . | S | . | S | R | . | S | . | S | S | . | S | S | . | R | . | S | ND | . | S | ND |
| **RF015GT** | + | . | S | . | S | . | S | R | S | . | S | . | S | . | S | S | . | S | . | S | S | . | S | S | . | S | . | S | ND | . | S | ND |
| **RF016SW** | ++ | S | S | S | S | S | S | S | S | R | R | S | S | S | S | R | S | S | S | S | S | S | S | S | S | S | S | S | ND | S | S | ND |
| **WH006AW** | ++ | R | R | S | S | S | S | R | R | S | S | R | R | R | R | R | R | R | S | S | S | S | S | S | S | S | R | . | ND | S | S | ND |
| **WH017KL** | + | S | S | S | S | S | S | S | S | S | S | S | S | S | S | R | S | S | S | S | S | S | S | S | S | S | S | S | ND | R | R | ND |
| **WH026NS** | Neg | ND | S | ND | S | ND | S | ND | S | ND | S | ND | S | ND | S | S | ND | S | ND | S | S | ND | S | S | ND | S | ND | S | ND | ND | R | ND |
| **WH036ES** | Neg | S | S | S | S | S | S | S | S | S | S | R | R | S | S | R | S | S | S | S | S | S | S | S | S | S | S | S | ND | S | S | ND |
| **WH037PD** | ++ | . | S | . | R | . | S | R | R | R | R | S | S | S | S | R | S | S | S | S | S | S | S | S | S | S | S | S | ND | S | S | ND |

"ND" = Sample or result not available. "." = Low sequence read coverage at position. "R" = Resistant, "S" = Susceptible

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

493 **Figure legends**

494 **Figure 1.** Comparison of the percentage of reference genome with at least one sequence read
495 covering a position on the y axis by the median depth of coverage on the x axis for each
496 individual sample. This is stratified by whether original sample material was a sputum or
497 MGIT culture.

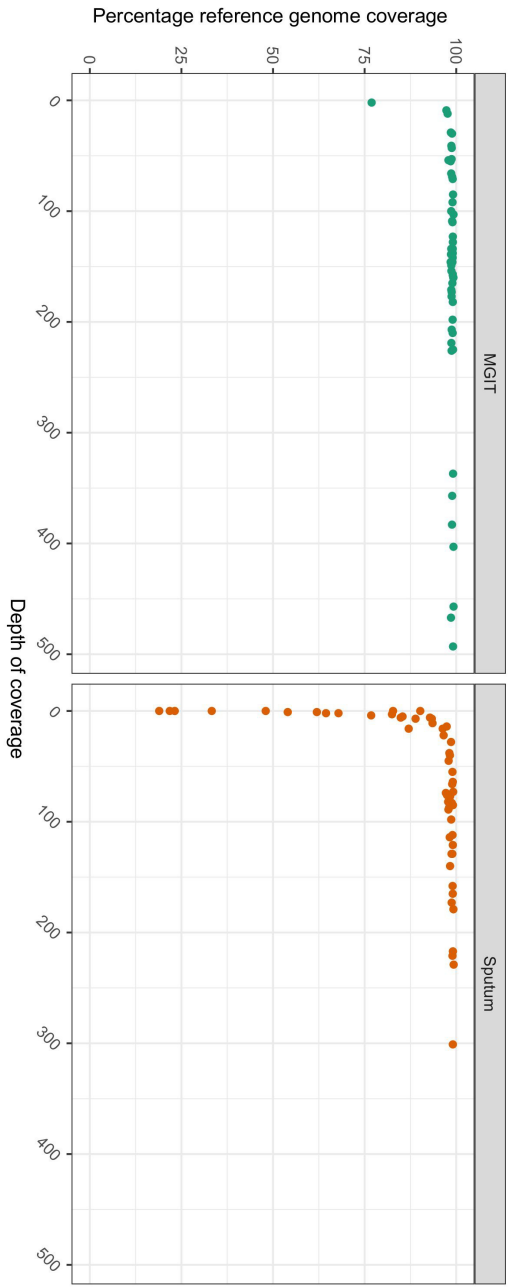
498 **Figure 2.** (A) Barplot showing the percentage reference genome coverage (A single read
499 covering each genome position) for patients with both a sputum and MGIT sample
500 sequenced. The plot is annotated with both Xpert MTB/RIF (GX) and smear microscopy
501 (Smear) results. For GX: **** = high, *** = medium, ** = low, and * = very low. For smear
502 *** = 3+, ** = 2+, * = 1+, S = scanty and - = negative. Where a result is missing the test was
503 not carried out. (B & C) Boxplot showing median depth of coverage for sputum samples
504 stratified by both the quantitative Xpert MTB/RIF measure (B) and semi-quantitative smear
505 result (C).

506 **Figure 3.** Heatmap clustering samples by the pairwise number of single nucleotide
507 differences between them. Sample names are formatted such that the patient identifier is at
508 the start followed by whether the sample originated from a MGIT culture or sputum.

509 **Figure 4.** The time taken in days on the x axis from sample collection (day 0) to when the
510 MGIT samples flag positive and sequence result becomes available is denoted by the grey
511 bars. The time taken for a patient sputum result to become available is marked by three
512 different identifiers depending on whether the sample underwent whole genome sequencing
513 or partial genome sequencing and whether the 24 hour or one hour hybridisation protocol was
514 used. Sputum samples that failed sequencing are marked by missing symbols. Patient
515 identifiers marked in red had confirmed drug resistant TB infections.



Journal of Clinical Microbiology



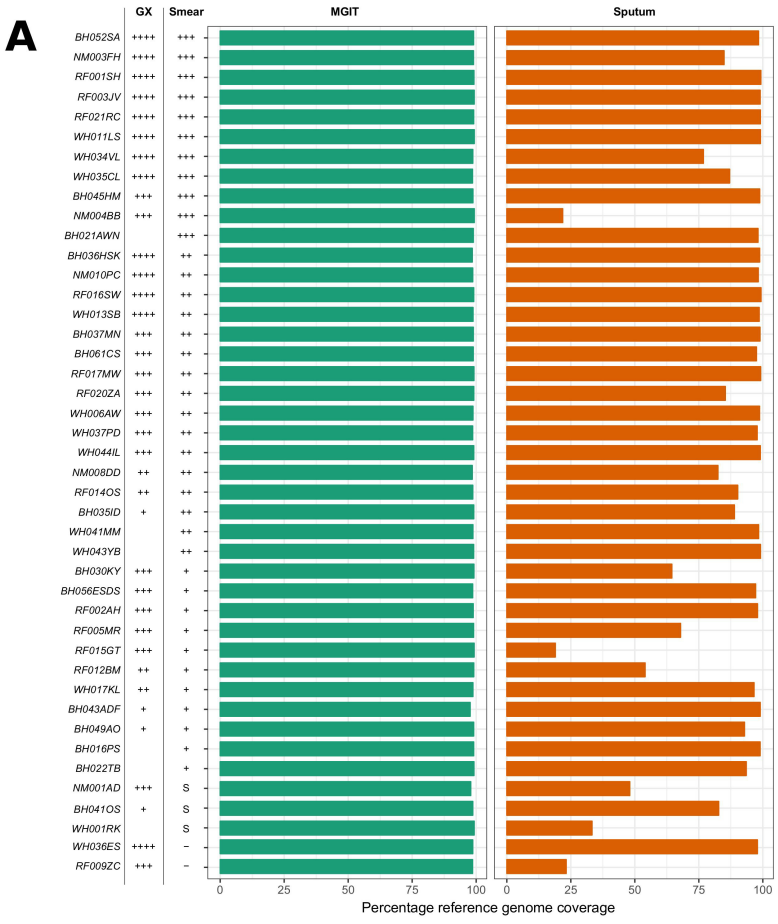

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

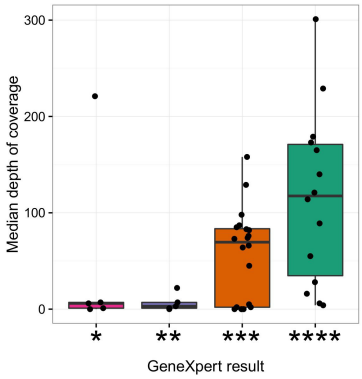
Accepted Manuscript Posted Online

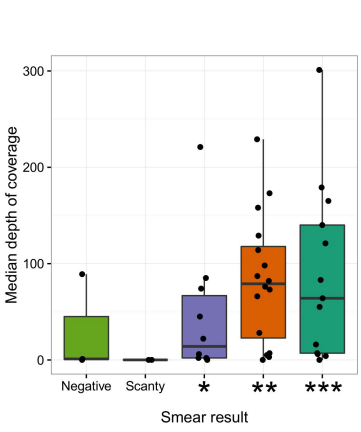
Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology




Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM



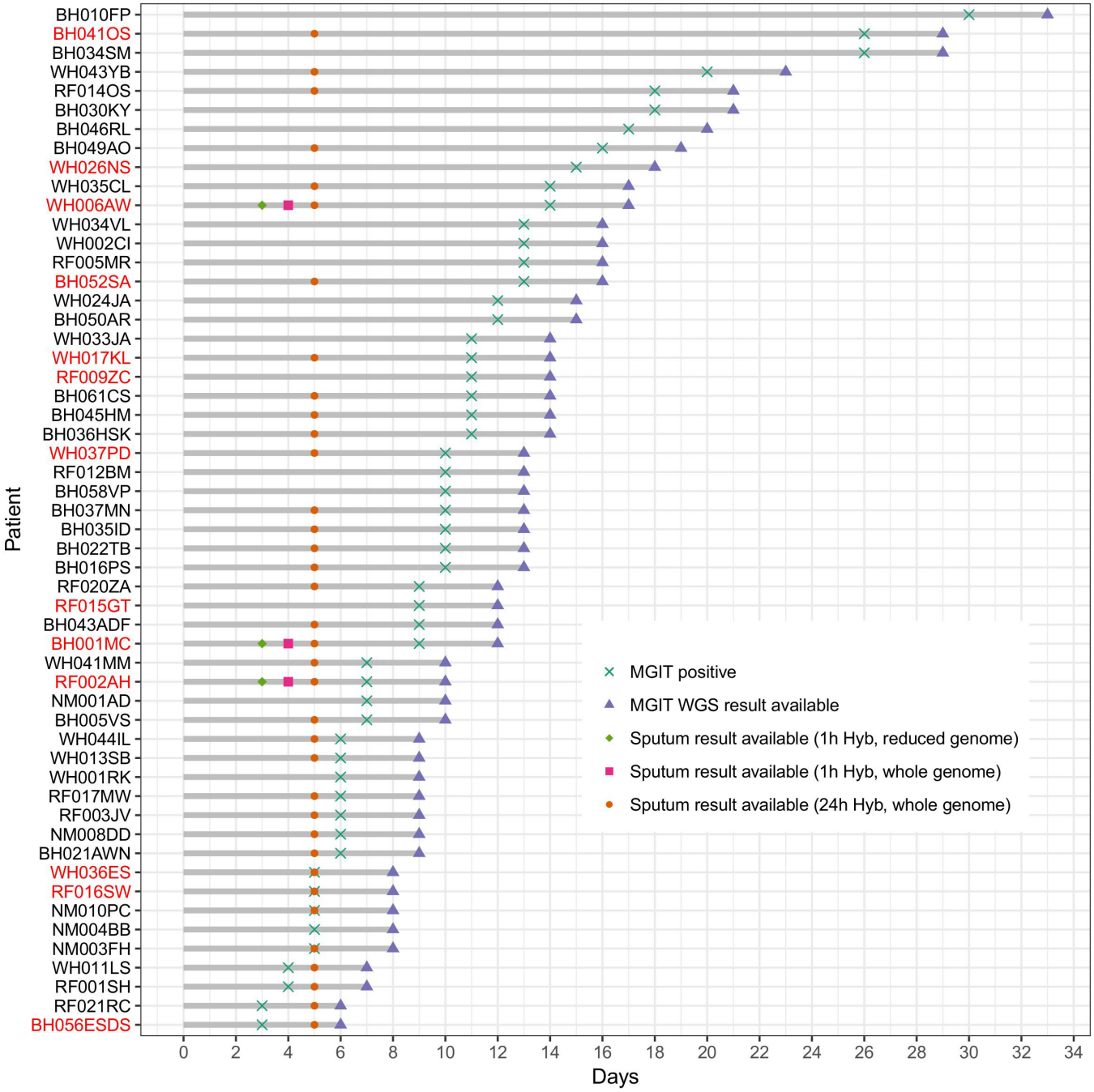

Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest

Accepted Manuscript Posted Online

Journal of Clinical Microbiology

JCM

Journal of Clinical Microbiology




Downloaded from http://jcm.asm.org/ on June 1, 2018 by guest