Improving diagnosis of tuberculosis from urine cell-free DNA

Amy Oreskovic

A dissertation
submitted in partial fulfillment of the
requirements for the degree of

Doctor of Philosophy

University of Washington
2021

Reading committee:
Barry Lutz, Chair
Paul Drain
Jr-Iuan Lai

Program Authorized to Offer Degree:
Bioengineering

University of Washington

Abstract

Improving diagnosis of tuberculosis from urine cell-free DNA

Amy Oreskovic

Chair of the Supervisory Committee:
Barry Lutz
Department of Bioengineering

Tuberculosis (TB) is the leading cause of infectious disease-related mortality worldwide, in part due to limitations in rapid diagnostics. Current TB tests rely on sputum samples, which are difficult to collect from many patients, including people living with HIV, those who are severely ill, and children, and may not detect extrapulmonary TB. Rapid sputum-based tests also have reduced sensitivity in these same patient populations, who more often have paucibacillary TB. There is a critical need for TB diagnostics that target more easily accessible samples, such as urine. Transrenal urine cell-free DNA (cfDNA) is a promising biomarker for TB but is challenging to detect because of the short length (<100 bp) and low concentration of TB-specific fragments. We aimed to increase the sensitivity of TB diagnosis from urine cfDNA by improving recovery of short fragments during the sample preparation step. Here, I present the development of a sequence-specific purification method for cfDNA with <1 copy/mL sensitivity from large volume (10 mL) urine samples that improves detection of short cfDNA relative to silica-based urine cfDNA extraction methods. In a clinical cohort study in South Africa, our assay had 84% sensitivity and 100% specificity for adult pulmonary TB, the highest diagnostic accuracy for a TB urine cfDNA test reported to date. We also characterized urine cfDNA using next-generation sequencing, revealing that TB-derived cfDNA is significantly more degraded than human-derived cfDNA. Finally, I explored several strategies for future assay improvements, including development of a multiplexed assay. The work described in this dissertation offers several important contributions to the growing evidence for urine cfDNA as a diagnostic target for TB, as well as to the broader cfDNA field. Collectively, it motivates continued

development of urine cfDNA-based assays for TB diagnosis and will aid informed design of improved cfDNA assays. Our sequence-specific purification approach for urine cfDNA has the potential to expand access to rapid TB diagnostics by enabling diagnosis across sputum-scarce and paucibacillary populations, addressing an urgent need that was identified as one of the highest priority gaps in TB diagnostics.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ACKNOWLEDGEMENTS

I would first like to thank my advisor, Barry Lutz. You've pushed me to become a better researcher, leader, and communicator, to be confident in myself and in my work, and to ask challenging questions. Your optimism, sense of humor, and enthusiasm for science inspire me. Thank you for giving me the freedom (and encouraging me) to take ownership of my project's direction. I would also like to thank the members of my supervisory committee, Paul Drain, James Lai, Wendy Thomas, James Carothers, and Jae-Hyun Chung, for their support and guidance throughout the course of my PhD.

The research presented in this dissertation would not have been possible without the contributions of many collaborators. I am extraordinarily appreciative to have collaborators committed not only to this project, but also to my personal development as a researcher. It has been a pleasure to work with and learn from you all. To Paul Drain, Adrienne Shapiro, and the PROVE-TB study team, thank you for the opportunity to conduct such exciting, clinically-relevant work. To David Horne, thank you for investing in this project in the earliest stages and helping it get off the ground. To Nuttada Panpradist, Diana Marangu, Norman Brault, and James Lai, thank you for your contributions to the initiation of this project and the opportunity to build upon your ideas. To Stephen Salipante, thank you for your eagerness to sign on to a new collaboration and helping explore exciting frontiers for TB urine cfDNA. I am also grateful to study participants for their contributions to my research and to the KwaZulu-Natal Department of Health and Edendale Hospital staff for their partnership.

I would like to acknowledge several sources of personal funding: the National Science Foundation Graduate Research Fellowship Program, the ARCS Foundation Scholar Award, and the UW Tuberculosis Research & Training Center Junior Investigator Award. Funding acknowledgements for research support follow each chapter.

Thank you to all the members of the Lutz Lab, past and present, for your advice and camaraderie. In particular, I would like to extend my gratitude to Nuttada Panpradist, Qin Wang, and Ian Hull, who have all at some point shared a sunless, closet-sized office with me and nevertheless made it an enjoyable place to work. I am thankful to have had such supportive lab mates with whom to share both the highs and lows of grad school.

Thank you to my family for your continual support and encouragement to dream big. To my Dad, thank you for fostering my lifelong sense of curiosity, inspiring my work ethic, and always catching my typos. To my Mom, thank you for modeling a spirit of relentless resilience and helping me to do the same. To my brother, thank you for helping me keep a healthy perspective – I hope to make you as proud as you make me.

Thank you to my friends, near and far. I am incredibly fortunate to have such a wonderful support system, and to have made many lasting friendships and fond memories over the past six years. I can't wait to find out what adventures await us next!

And finally, to my partner Seth, thank you for your love, support, and patience every step of the way. There are not enough words in this dissertation to explain how thankful I am to have you in my life. I could not have done this without you and am so excited to see what the future holds for our lives together.

1. INTRODUCTION

1.1. SIGNIFICANCE OF PROBLEM

With an estimated 10 million cases and 1.4 million deaths in 2019, tuberculosis (TB) is the leading cause of global mortality due to infectious disease (1). Despite widely available treatment and a high treatment success rate, insufficiencies in rapid, sensitive TB diagnosis delay or prevent appropriate treatment, contributing to increased disease transmission and mortality. Current TB diagnostics rely on the collection of sputum samples, which may be difficult to obtain and/or have reduced sensitivity in key patient populations including those with HIV-associated TB, pediatric TB, and extrapulmonary TB (EPTB). There is an urgent need for new TB diagnostics that target alternate sample types, such as urine, with high sensitivity across patient populations (2). Urine contains short (30–100 bp), dilute (<1–200 ng/mL) cell-free DNA (cfDNA) fragments (3), including a small number of TB-specific fragments that can be used to detect active TB infection (median approximately 10 copies/mL) (4–6). Unfortunately, recovery by conventional silica-based DNA extraction decreases with decreasing fragment length and concentration and is thus inadequate for urine cfDNA. Previous studies targeting urine cfDNA for TB diagnosis have used sample preparation and/or amplification methods not suitable for short fragments, leading to low, variable clinical sensitivities.

1.2. PROPOSED SOLUTION

We hypothesized that we could increase the sensitivity of TB diagnosis from urine cfDNA by 1) improving the recovery of short, dilute urine cfDNA fragments during sample preparation and 2) improving designs for specific amplification of short target sequences. We identified sequence-specific purification as a method likely to perform well for urine cfDNA because it is agnostic to both fragment length and concentration. The primary goal of my thesis project was to develop a maximum sensitivity sequence-specific purification and short-target polymerase chain reaction (PCR) assay and demonstrate its ability to detect TB cfDNA in clinical urine samples. A graphic abstract is shown in Figure 1.1.

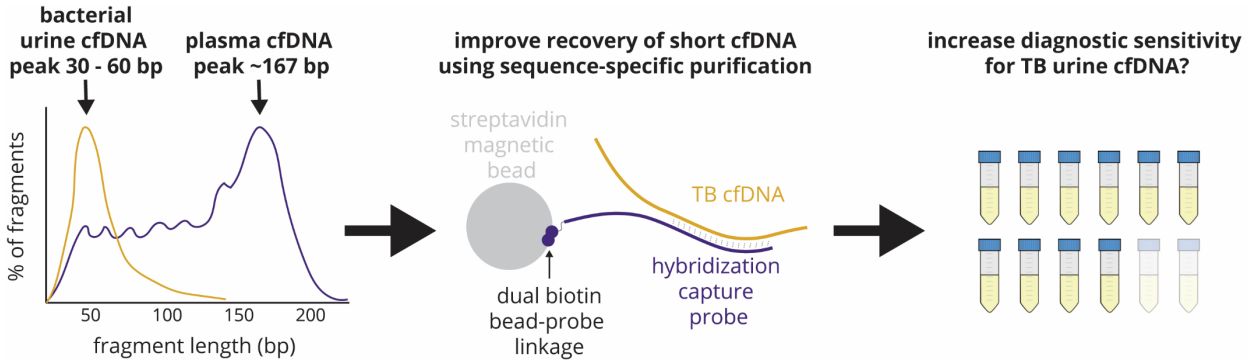


Figure 1.1: Graphic abstract. *We hypothesized that improving recovery of short urine cfDNA using sequence-specific purification would increase the sensitivity of TB diagnosis from urine cfDNA.*

1.3. SUMMARY OF DISSERTATION WORK

1.3.1. *Development of sequence-specific purification method for urine cell-free DNA*

My first objective was to develop a sequence-specific purification method that uses oligonucleotide capture probes immobilized on magnetic beads to improve extraction of short cfDNA from large-volume urine samples. The benchmark for success was sufficient sensitivity for detection of TB-specific cfDNA in urine. Due to the low concentrations of TB cfDNA in urine, I had to invest significant effort into optimizing the assay's analytical performance to achieve this aim. Relative to previous implementations of sequence-specific purification for sample preparation, I increased the input sample volume (from ≤1 mL to 10 ml) and pushed the limits of analytical sensitivity (to achieve a limit of detection of ≤1 copy/mL). In addition, I developed strategies for PCR amplification of short targets with maximal specificity. As the output of this work, I presented a detailed, user-ready protocol for sequence-specific purification of urine cfDNA (with guidelines for designing new capture probes for applications beyond TB) and characterized the analytical performance of our TB urine cfDNA assay. I also broadly described the assay development process and key design features that led to its strong performance.

1.3.2. *Analytical comparison of urine cell-free DNA extraction methods*

My second goal was to compare the analytical performance of sequence-specific purification with alternate urine cfDNA extraction methods. The scope of this work was intended to go beyond TB and be useful to a broader audience interested in the diverse applications of urine cfDNA, although one of the methods tested (Wizard/GuSCN) has been used for TB urine cfDNA (4). The aim of this work was to better understand the current options for urine cfDNA extraction and develop a representative dataset that could be used to guide selection and optimization of extraction methods for urine cfDNA analysis. For each extraction method, I characterized the dependence on DNA fragment length, performance at

low DNA concentrations, tolerance to variable urine conditions, and susceptibility to PCR inhibition. This work will enable informed selection of sample preparation methods, which may lead to increased clinical sensitivity and reproducibility of urine cfDNA diagnostics. A secondary benefit of this work was that it confirmed improvement in recovery of short cfDNA using sequence-specific purification relative to other urine cfDNA extraction methods, particularly the Wizard/GuSCN method used previously for TB urine cfDNA, supporting our hypothesis that sequence-specific purification has the potential to increase the diagnostic sensitivity of urine cfDNA for TB.

1.3.3. *Detection of tuberculosis cell-free DNA in clinical urine samples*

My next aim was to determine the diagnostic accuracy of the TB urine cfDNA assay in clinical urine specimens. In a clinical cohort study, we analyzed urine from 79 participants: 49 TB-positive participants in South Africa, 10 TB-negative participants in South Africa, and 15 healthy controls in the United States. In addition to calculating assay sensitivity and specificity, we compared assay performance across subgroups (HIV status, CD4 count, smear status, urine LAM status) and analyzed detected cfDNA concentrations for correlations with clinical data (CD4 count, days to culture positivity). The results of this study showed that our approach is promising for the diagnosis of TB from urine cfDNA and contributed to the growing body of evidence for urine cfDNA as a biomarker for TB.

1.3.4. *Next-generation sequencing of tuberculosis urine cell-free DNA*

Urine cfDNA has not been well characterized as a biomarker for TB. We aimed to investigate the size and composition of TB-derived urine cfDNA using next-generation-sequencing (NGS). To minimize bias during analysis of highly fragmented urine cfDNA, we selected a combination of DNA extraction (Q Sepharose) and single-stranded library preparation (SRSLY) methods demonstrated to recover short cfDNA fragments. We sequenced cfDNA from 12 urine samples, compared the distributions of TB-derived and human-derived cfDNA, and examined cfDNA coverage of the TB genome for over-represented regions that could serve as new diagnostic targets. This work provides critical validation needed for urine cfDNA as a biomarker for TB and will inform future design of TB urine cfDNA assays.

1.3.5. *Next steps towards continued assay development*

In the final portion of my dissertation work, I sought to explore several strategies for future assay development. Although both the analytical and clinical performance of the TB urine cfDNA were promising, the low concentrations of TB-specific cfDNA detected in urine presented an opportunity to increase sensitivity by improving the detection of low concentration samples. I designed two approaches

to expand the fraction of detectable cfDNA molecules – multiplexing and ultrashort PCR. I also investigated two additional strategies that may improve reliability of detection of the current cfDNA target sequence – increasing the input sample volume and adding a proteinase K digestion step to mitigate the effects of sample variability. I aimed to take the first steps towards implementation of each strategy, identify which strategies were most promising, and outline the recommended next steps for future assay development.

1.3.6. *Summary*

The results of my dissertation work illustrate the advantages of our sequence-specific approach and demonstrate that, with improved sample preparation, urine cfDNA is a viable diagnostic target for TB. In addition to the development of a sensitive TB urine cfDNA assay, this work also offers several important contributions to the cfDNA field. Specifically, it substantially adds to the growing evidence for urine cfDNA as an approach to TB diagnostics and will help inform design of better urine cfDNA assays, including those for applications besides TB. Future work, beyond the scope of this project, will aim to simplify our TB urine cfDNA assay into a format more suitable for use in resource-limited settings.

2. BACKGROUND

Detailed background about TB, urine cfDNA, and sequence-specific purification is given in this chapter, with relevant background summarized again at the beginning of each chapter to follow. Note that several chapters are adapted from published papers. For completeness, I left the introductions from published papers relatively unmodified, so there may be some minor repetition with the background information presented here.

2.1. NEED FOR IMPROVED TB DIAGNOSTICS

2.1.1. *Overview*

Tuberculosis (TB) is the leading cause of global mortality due to infectious disease, with an estimated 10 million cases and 1.4 million deaths in 2019 (1). The TB disease burden is carried disproportionately by developing countries, with 87% of cases located in just 30 high-burden countries (1). With proper treatment, TB is largely curable (85% treatment success rate), yet an estimated 30% of TB cases remain undiagnosed or unreported, in part due to limitations in rapid diagnostics (1). Insufficient access to rapid, sensitive TB diagnostics delays or prevents appropriate treatment, contributing to increased mortality, increased disease transmission, and development of drug resistance. Current TB tests rely on

sputum samples, which are difficult to collect from people living with HIV, severely ill patients, and children, and may not detect extrapulmonary TB (EPTB). Rapid sputum-based tests (e.g., smear microscopy, Xpert MTB/RIF) also have lower sensitivity in these same underserved patient populations, who more often have paucibacillary TB (7–9). A World Health Organization (WHO) consensus meeting identified a rapid, non-sputum-based test as one of the highest priority target products for TB diagnostics (2), emphasizing the critical need for new TB diagnostics targeting alternate sample types with high sensitivity across patient populations.

2.1.2. *Sputum-based TB diagnostics – current gold standard and limitations*

The gold standard for TB diagnosis, sputum culture, is not available in most resource-limited settings and requires several weeks for growth. Sputum smear microscopy, where a sputum sample is spread on a microscope slide, stained, and counted for bacteria, is more common in resource-limited settings. Smear microscopy is highly user-dependent, has variable and limited sensitivity (20 – 80%) (10), and often requires sputum collection across multiple days. There are now several rapid nucleic acid amplification tests (NAAT) recommended by the WHO, the most commonly used of which are the Cepheid Xpert MTB/RIF and Xpert MTB/RIF Ultra. These are both cartridge-based tests run on Cepheid's GeneXpert system that simultaneously detect TB and screen for resistance to rifampin, one of the first line TB drugs, within two hours. Xpert MTB/RIF was widely implemented starting in 2010. With an overall pooled sensitivity of 85% (7), it offers a substantial increase in sensitivity over smear microscopy but still has suboptimal sensitivity compared to culture. In 2017, the next-generation Xpert MTB/RIF Ultra cartridges were recommended as a replacement with higher sensitivity than Xpert MTB/RIF cartridges. In a head to head comparison of Xpert MTB/RIF Ultra and Xpert MTB/RIF in the same panel of sputum specimens, they had sensitivities of 88% and 83% and specificities of 96% and 98%, respectively (11). Because both tests rely on an expensive external instrument, they are not suitable for use outside centralized settings. Cepheid has a portable near-patient device, the GeneXpert Omni system, in development, but it has not yet been endorsed by the WHO.

The limitations of current sputum-based diagnostics are most readily apparent for smear-negative or paucibacillary TB, where there is a lower bacterial load in the lungs. This is more common in patients with HIV-associated TB, pediatric TB, and extrapulmonary TB. HIV-TB co-infection accounts for 8% of TB cases and contributes disproportionately to TB mortality (208,000 deaths annually) (1). HIV infection increases the risk of developing active TB disease, and HIV-TB co-infection is one of the leading causes of death in people living with HIV (32% of HIV-related deaths) (12). Pediatric TB and EPTB make up an

estimated 12% and 16% of TB cases, respectively (1). In addition to limited sputum sample availability in some of these key underserved patient populations, Xpert MTB/RIF has reduced sensitivity. For smear-negative TB, the pooled sensitivity of Xpert MTB/RIF drops to 67%, compared to 98% for smear-positive TB (7). For HIV-positive patients compared to HIV-negative patients, Xpert MTB/RIF pooled sensitivity decreases from 89% to 77% (7). In pediatric patients with pulmonary TB, Xpert MTB/RIF pooled sensitivity is 62% (8). Xpert MTB/RIF Ultra helps lessen the reduction in sensitivity in these populations, but still does not rival culture. In a direct comparison on a panel of smear-negative specimens, Xpert MTB/RIF Ultra had sensitivity of 63% compared to 46% for Xpert MTB/RIF (11). In the same study, sensitivity was 90% for Xpert MTB/RIF Ultra versus 77% for Xpert MTB/RIF in people living with HIV (11). In children, Xpert MTB/RIF Ultra had 74% sensitivity (13). Sensitivity for EPTB diagnosis using both Xpert MTB/RIF and Xpert MTB/RIF Ultra varies, and is generally higher if the sample tested corresponds to the site of infection rather than sputum (9, 14).

Availability of sputum samples is also a limitation for existing rapid tests. Over a third of hospitalized patients with HIV are estimated to be sputum-scarce, or have difficulty providing a sputum sample (15). Collecting adequate respiratory samples from children is challenging, frequently requiring invasive sputum induction or collection of gastric aspirates, which cannot be readily obtained in low resource settings and often require hospitalization of the child (16). Finally, extrapulmonary TB often requires collection of a sample from the site of infection, rather than a sputum sample.

2.1.3. *Non-sputum-based TB diagnostics*

The WHO has published a target product profile highlighting the need for "a point-of-care non-sputum-based test capable of detecting all forms of TB" and outlining the optimal requirements for pulmonary TB in adults as ≥98% sensitivity for smear-positive TB, ≥68% sensitivity for smear-negative TB, and ≥80% sensitivity for HIV-associated TB (2). Moving away from the reliance on sputum samples will not only increase sample availability but will ideally also increase diagnostic sensitivity for HIV-associated TB, pediatric TB, and EPTB. Urine is an attractive alternate sample for TB diagnosis because it is easy to collect and poses minimal risk of TB disease transmission. Xpert MTB/RIF has low sensitivity in urine for pulmonary TB diagnosis (4–48%) (17–19). The Alere Determine TB LAM is a commercially available lateral flow test targeting lipoarabinomannan (LAM), a lipopolysaccharide in mycobacterial cell walls that is excreted in the urine of individuals with active TB. Sensitivity of urine LAM detection is limited, especially in HIV-negative individuals. Sensitivity is inversely proportional to CD4 cell count, with a pooled sensitivity in HIV-positive individuals of 42% for Alere Determine TB LAM (20). The WHO only

recommends the use of Determine TB LAM for people living with HIV and does not recommend it as a general screening test for TB (21). Ongoing development of several next-generation urine LAM assays aims to improve sensitivity through concentration and/or signal amplification steps. One such test furthest along in development is the Fujifilm SILVAMP TB LAM, which improves sensitivity relative to Alere Determine TB LAM in both HIV-positive (70% vs. 42%) and HIV-negative (53% vs. 11%) individuals (22, 23). While these advances are encouraging, as antigen-based tests, LAM assays may not be able to achieve the same sensitivity as NAATs. Conversely, they may be advantageous over NAATs due to their simplicity and ease of use afforded by the lack of an amplification step.

Urine from patients with active TB also contains cell-free DNA (cfDNA) fragments specific to TB (4). Urine cfDNA has not yet been well-characterized as a diagnostic target for TB, but has been previously detected with similar sensitivities in both HIV-negative and HIV-positive patients (4–6). As NAATs, urine cfDNA assays enable amplification and the possibly of increased sensitivity across patient populations. The focus of my dissertation research was to develop improved sample preparation and amplification methods for TB diagnosis from urine cfDNA. Additional background information for urine cfDNA is given in the following section.

In addition to urine, oral swabs and stool have been investigated as alternative samples to sputum. In South Africa, oral swab analysis had 83% sensitivity and 92% specificity in adults (24) and 43% sensitivity and 93% specificity in children (25). In Peru, an oral swab assay from a different research group had 51% sensitivity and 97% specificity in adults (26) and 21% sensitivity and 99% specificity in children (27). In a pilot study combining oral swab samples with Xpert MTB/RIF Ultra, the sensitivity and specificity were 45% and 100%, respectively (28). Stool PCR has been most commonly applied for diagnosis of pediatric TB, with pooled sensitivity and specificity of 57% and 98%, respectively (29). While stool PCR has been applied less frequently for adults, a recent study showed 97% sensitivity for smear-positive TB, 77% sensitivity for smear-negative TB, 69% sensitivity for EPTB, and 95% overall specificity (30).

2.2. Urine cell-free DNA as a biomarker for TB

2.2.1. *Origins, characterization, and clinical applications*

Circulating cfDNA was first described in 1948 by Mandel and Metais, with approximately 4.7 ng/mL of DNA extracted from plasma of 10 healthy patients (31). In the same study, an additional 15 patients with conditions including pregnancy, autoimmune disorder, and tuberculosis were investigated, with elevated levels of cfDNA observed in some cases (31). Further observations of elevated plasma cfDNA

concentrations in cancer patients later spurred interest in the potential of cfDNA as a diagnostic target (32). In 1994, two separate groups reported the presence of specific oncogene mutations in the plasma cfDNA of patients with acute myelogenous leukemia (33) and pancreatic cancer (34). Additional studies demonstrating the diverse origins of circulating cfDNA soon followed. Y chromosome cfDNA was detected in plasma of 80% (24/30) of women pregnant with male fetuses, but not in plasma of women pregnant with female fetuses (0/13) (35). Y chromosome DNA was observed in only 17% of samples from nucleated blood cells, confirming the cell-free nature of the detected DNA. Y chromosome cfDNA was also detected in the plasma of 100% (6/6) and 82% (14/17) of women who had received liver and kidney transplants, respectively, from a male donors (36). Circulating cfDNA arises through a variety of mechanisms. The bulk of cfDNA is generated by breakdown of dying cells (via apoptosis, necrosis, pyroptosis, phagocytosis, etc.), but cfDNA can also be generated by active release mechanisms (37). Aucamp *et al*. provide a thorough review of the diverse mechanisms by which cfDNA is generated (37).

In 2000, Botezatu *et al.* discovered that a fraction of circulating cfDNA, composed of short fragments, crosses the kidney barrier and is excreted in urine (3). To demonstrate the circulatory origins of this urine cfDNA, they subcutaneously injected human DNA into mice and were able to amplify human-specific sequences from urine. They were able to detect male-specific sequences in the cell-free urine of 56% (5/9) of women transfused with male blood and 80% (8/10) of women pregnant with male fetuses. They also detected *K-RAS* mutations in the cell-free urine of 80% (4/5) of patients with corresponding tumor mutations. Urine cfDNA is now an established or emerging noninvasive biomarker for cancer mutation detection (3, 38, 39), infectious disease diagnosis (4, 40–42), organ transplant monitoring (43, 44), and prenatal genetic screening (3, 45, 46). The subfraction of urine cfDNA derived exclusively from circulating cfDNA is known as transrenal DNA (trDNA), but cfDNA can also be generated directly in urine from cells shed along the urinary tract.

The concentration range reported for total urine cfDNA is <1 ng/mL to 200 ng/mL, similar to that for plasma DNA (3, 47, 48). While plasma cfDNA of human origin has a peak length of 160–167 bp, reflecting protection of histone-associated DNA within nucleosomes (45, 49–51), urine cfDNA is more fragmented (Figure 2.1A). The upper limit of the transrenal fraction of urine cfDNA is defined by glomerular filtration (~70 kDa, approximately 100 bp) (52) and cfDNA fragments are rapidly degraded further in urine (53, 54). The transrenal fraction of urine cfDNA is expected to be largely unassociated with proteins, which are typically unable to cross the kidney barrier. Faster degradation kinetics for naked DNA compared to protein-associated DNA, combined with faster degradation kinetics of DNA in urine compared to blood,

result in highly fragmented urine cfDNA (53, 54). Determining the true fragment length distribution of urine cfDNA is challenging because both DNA extraction and sequencing library preparation methods may underestimate the prevalence of shorter fragments (49, 51), but the majority of fragments are expected to be <100 bp (45, 54–56), with a wider distribution of fragments around the peak fragment length compared to plasma cfDNA (56). cfDNA fragment length distribution varies across patients, with specific diseases and physiological conditions (e.g., infectious disease, immunosuppression, cancer, pregnancy), and depending on the origin of the cfDNA. Non-nucleosomal urine cfDNA, including bacterial, viral, and mitochondrial cfDNA, are more fragmented than human chromosomal cfDNA because they are less protected by DNA-associated proteins while in circulation (Figure 2.1B-C) (45, 49, 53, 55). For these particularly short forms of urine cfDNA, the peak fragment length may be as short as 30–60 bp (45, 49, 55). Fetal- (45) and tumor-derived (57) cfDNA are also more degraded than maternal and non-tumor cfDNA, respectively. Recently, new single-stranded library preparation methods for next-generation sequencing have revealed that very short, formerly undetectable fragments make up a larger fraction of both plasma (49, 51, 58) and urine (55) cfDNA than previously recognized. In plasma, single-stranded library preparation showed that nearly half of cfDNA was <120 bp (mean 62–91 bp) and was undetectable by conventional library preparation (58). The fragment length distribution of TB-derived urine cfDNA has not yet been determined but is assumed to consist largely of short fragments.

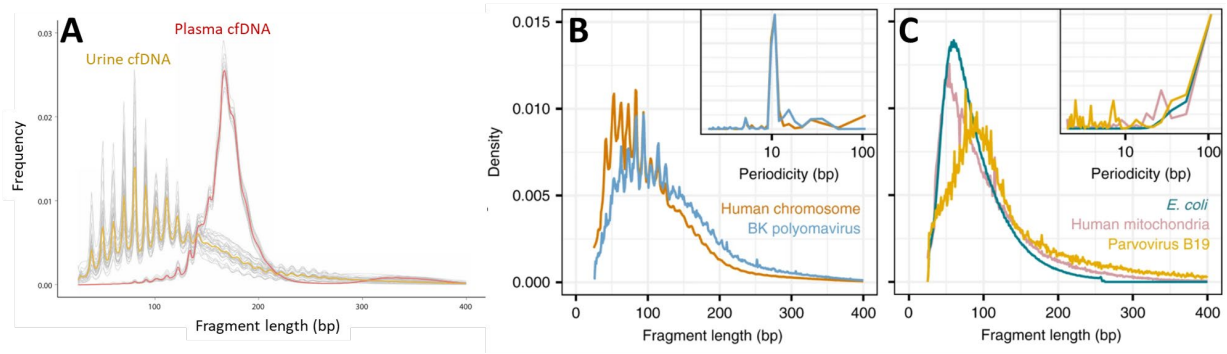


Figure 2.1: Examples of urine cfDNA fragment length distribution. (A) *Urine cfDNA is shorter than plasma cfDNA. Image adapted from Markus et al. (2021)* (56) *with permission from AAAS.* ***(B & C)*** *Single-stranded library preparation reveals that urine cfDNA derived from non-nucleosomal sources (bacteria, mitochondria, viruses) is more fragmented than chromosomal urine cfDNA. Image reprinted from Burnham et al. (2018) (55) under a CC BY 4.0 license.*

2.2.2. *Urine cfDNA for TB diagnosis*

As TB bacteria in the lungs or elsewhere throughout the body die, they generate circulating cfDNA specific to TB. A fraction of this circulating TB cfDNA crosses the kidney barrier and can potentially be

used to diagnose active TB infection (Figure 2.2). Previous studies targeting urine cfDNA for pulmonary TB diagnosis are summarized in Table 2.1. Cannas *et al.* were the first to definitively show the presence of *cell-free* TB DNA in urine (4). Up until that point, and in some studies since, the cell-free nature and/or extremely short length of the target were not widely recognized, and methods were not designed appropriately for short, dilute cfDNA fragments. See Section 2.2.3 for an in-depth discussion of these limitations. Briefly, these studies performed a centrifugation step, discarded the soluble fraction, extracted DNA from the sediment, and amplified long DNA targets (≥192 bp, many >400 bp), thus likely detecting high molecular weight cell-associated DNA rather than cfDNA. Included in these studies were three that applied the Xpert MTB/RIF assay (192 bp target) to urine, with low sensitivity (4–48%) (17–19). We identified only six studies that retained the soluble urine fraction for at least some of the samples tested. One of these studies (Bordelon *et al.*) used a DNA extraction method shown to have low recovery of short fragments (59). Another (Peter *et al.*) used the Xpert MTB/RIF assay, which has a relatively long amplicon length of 192 bp (17). Yet another (Fortún *et al.*) used the Hologic Amplified Mycobacterium Direct Test (MDT), with unspecified amplicon length (60). The three remaining studies (Cannas *et al.*, Labugger *et al.*, and Patel *et al.*) are the only ones to confidently detect *cell-free* TB DNA in urine with methods most appropriate for short fragments (4–6).

When considering all studies, the overall sensitivity of TB cfDNA detection is low and variable, likely due to nonoptimal and inconsistent sample preparation and amplification methods. Limiting to the three studies definitively detecting cfDNA with the most appropriate methods for short fragments, the sensitivity is still limited (40–79%) (4–6). In addition to the urine cfDNA studies listed here, Click *et al.,* Ushio *et al.,* and Lyu *et al*. also detected TB-specific cfDNA in the plasma of patients diagnosed with pulmonary TB with sensitivities of 45%, 65%, and 41%, respectively (61, 62).

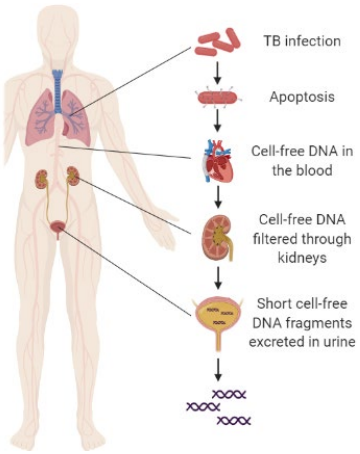


Figure 2.2: Origins of TB urine cfDNA. *Original image modeled after Green et al.* (52).

Table 2.1: Previous studies targeting urine cfDNA for pulmonary TB diagnosis. *Studies with sample preparation and amplification methods designed most appropriately for short, dilute urine cfDNA are bolded. For a simplified table highlighting more details about these particular studies, see Table 5.5 in Chapter 4.8. This literature review builds upon a Global WACh capstone project by Nuttada Panpradist and Dr. Diana Marangu, which led to the initiaion of the TB urine cfDNA project described here.*

First author (year)	Sample preparation			PCR detection		Sensitivity		Specificity
	Appropriate for short urine cfDNA?	Retention of urine supernatant?	Extraction method	Designed for short urine cfDNA? (amplicon length)[a]	Genomic target	HIV-negative	HIV-positive	
Sechi (1997) (63)	No	No	Lysis only (no purification)	No (567/108 bp)	IS6110	6.3% (12/190)	15.7% (65/412)	Not tested
Aceti (1999) (64)	No	No	Lysis only (no purification)	No (567/310 bp)	IS6110	Not tested	100% (13/13)	99.3% (142/143)
Kafwabulula (2002) (65)	No	No	Phenol-chloroform	No (567/166 bp or 567/108 bp)	IS6110	23% (3/13)	64% (32/50)	98% (62/63)
Torrea. (2005) (66)	No	No	Phenol-chloroform	No (567/310 bp)	IS6110	32% (42/131)	53% (46/86)	98% (54/55)
Rebollo (2006) (67)	No	No	Lysis only (no purification)	Moderately (123 bp)	IS6110	8% (2/24)	37% (7/19)	100% (26/26)
Cannas (2008) (4)	**Yes (claimed)**	**Yes**	**Wizard silica resin/GuSCN**	**Moderately (129/67 bp)**	**IS6110**	**79% (34/43)**	**Not tested**	**100% (23/23)**
Gopinath (2008) (68)	No	No	Phenol-chloroform	No (786 bp)	Cfp32	52% (24/46, HIV status not specified)		100% (121/121)
da Cruz (2011) (69)	No	No	Glass matrix	No (409/316 bp)	IS6110	30% (10/33)	Not tested	91% (89/98)
Amin (2011) (70)	No	No	Ethanol precipitation kit	Moderately (123 bp)	IS6110	47% (105/225, HIV status not specified)		Not tested
Peter (2012) (17)	No	Mostly no; yes for a smaller number of samples	Xpert cartridge	No (192 bp)	rpoB	Not tested	48% concentrated urine (54/113); 8% unconcentrated urine (3/38)	98% (61/62)
Shenai (2013) (18)	No	No	Xpert cartridge	No (192 bp)	rpoB	4% (1/26)	Not tested	Not tested
Lawn (2013) (19)	No	No	Xpert cartridge	No (192 bp)	rpoB	Not tested	19% (16/84)	Not tested
Fortún (2014) (60)	Maybe	Yes	Lysis only (no purification)	Unspecified (Hologic Amplified MDT Test)	16S rRNA	18% (5/28, HIV status not specified)		Not tested
da Costa Lima (2015) (71)	No	No	Glass matrix	No (409/316 bp)	IS6110	35% (14/40)	Not tested	100% (15/15)
Labugger (2016) (5)	**Yes (claimed)**	**Yes**	**Unspecified silica resin (protocol similar to Wizard/GuSCN)**	**Yes (38 bp)**	**IS6110**	**64% (7/11)**	**Not tested**	**100% (8/8)**
Bordelon (2017) (59)	No	Yes	Dynabeads MyOne Silane (low recovery)	Yes (67 bp)	IS6110	0% (0/33)	0% (0/18)	0% (0/31)

Patel. (2018) (6)	**Yes (claimed)**	**Yes**	**Unspecified silica resin (protocol similar to Wizard/GuSCN)**	**Yes (38 bp)**	**Direct repeat (DR) region**	**40% (36/90)**	**45% (38/84)**	**89% (210/237)**

[a] Amplicon length for studies that used nested PCR are given as (outer amplicon length/inner amplicon length). Note that while nested PCR can increase sensitivity compared to unnested PCR, all four primer sites must be located on a single DNA fragment for successful amplification.

2.2.3. *Sample preparation and amplification challenges for urine cfDNA*

The low and variable sensitivities reported in previous TB urine cfDNA studies are likely due to a combination of sample preparation methods that are inefficient at capturing short urine cfDNA fragments and PCR detection methods inappropriately designed for short, dilute urine cfDNA fragments. For high sensitivity detection of TB urine cfDNA, there are three critical methodological requirements, as illustrated in Figure 2.3: (1) retention of the urine supernatant, (2) DNA extraction with high recovery of short fragments, and (3) PCR amplification targeting short fragments.

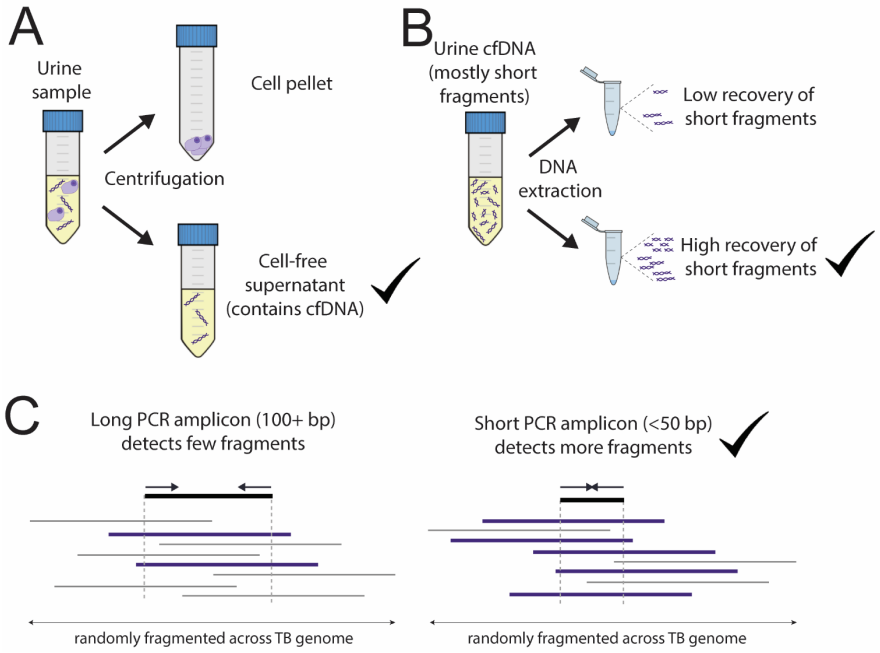


Figure 2.3: Key criteria for sample preparation and amplification methods targeting urine cfDNA. (A) *Cell-free DNA is contained in the soluble fraction. If urine is centrifuged prior to extraction, cfDNA will be in the supernatant, which should be retained. Discarding urine supernatant and analyzing the cell pellet will miss most cfDNA.* ***(B)*** *DNA extraction method must have high recovery of short fragments, which make up most urine cfDNA.* ***(C)*** *PCR must target a short amplicon. Urine cfDNA is randomly fragmented across the genome. Both primer sites must be contained on an intact fragment for successful amplification. Short PCR will successfully amplify a larger fraction of degraded urine cfDNA fragments.*

2.2.3.1. Retention of urine supernatant

The soluble fraction (supernatant) of the urine sample, which contains cfDNA, must be retained and analyzed (Figure 2.3A). Because the cell-free nature of TB DNA in urine was not initially recognized, many studies included a centrifugation step to concentrate larger volumes of urine prior to extraction (18, 19, 72, 63, 64, 66–71). By discarding the supernatant and analyzing only the cell pellet, these studies lost most, if not all, cfDNA. Comparison of matched supernatants and pellets revealed detection of TB DNA in 7/8 supernatants compared to 2/8 pellets (4). An appropriately designed method for urine cfDNA would centrifuge and analyze the cell-free supernatant, or analyze whole urine. Only six TB urine cfDNA studies that analyzed whole urine or the cell-free supernatant were identified (4–6, 17, 59, 60).

2.2.3.2. DNA extraction with high recovery of short fragments

The second key aspect for sample preparation of urine cfDNA is the selection of a DNA extraction method capable of recovering short, dilute cfDNA fragments (Figure 2.3B). Unfortunately, conventional silica-based extraction methods for cell-associated DNA or even plasma cfDNA are not suitable for urine cfDNA because recovery decreases with decreasing fragment length and concentration. The “Boom” method, common in both commercially-available kits and lab-based protocols, adsorbs DNA to silica under chaotropic conditions (73). The driving forces of silica adsorption are hydrophobic interactions due to dehydration of silica and DNA surfaces and hydrogen bonding between silica and the DNA backbone, both of which depend on DNA length (74–76). As a result, silica adsorption is less effective at purifying short fragments. The lower length limit of silica-based methods varies across binding matrices and buffer conditions, but recovery generally decreased below 50–100 bp (Figure 2.4). Silica adsorption also requires relatively high DNA concentrations for optimal performance since a fraction of DNA may remain irretrievably bound to the silica surface (75, 76). This loss is trivial in most samples, but for low concentration samples, like urine cfDNA, may make up a significant portion of the input (75).

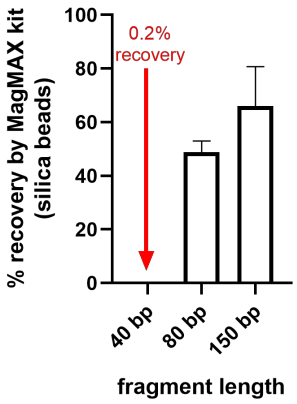


Figure 2.4: Silica adsorption is dependent on DNA fragment length. *Silica adsorption has poor recovery of short DNA fragments. As an example, results of purification of 10^5 copies/mL of spiked DNA using the Thermo Fisher MagMAX Cell-free DNA Isolation Kit are given (mean ± SD, n=3).*

Cannas *et al.* used a silica resin-based extraction method designed specifically for urine cfDNA, which they claim improves recovery of short fragments (4). The method adsorbs cfDNA to Wizard silica resin in the presence of an especially high concentration of guanidinium chaotropic salt (>3M), and is the same as the method used by Botezatu *et al.* to originally demonstrate the presence of cfDNA in urine (3). Labugger *et al.* and Patel *et al.* both used a similar silica resin-based method, but did not specify the type of resin or the binding conditions (5, 6). Bordelon *et al.* observed no difference in TB cfDNA concentration in patients with and without TB, likely due to choice of extraction method (59). Previous work by the same group showed that their extraction method (Dynabeads MyOne Silane magnetic beads within a surface tension valve device) had lower recovery of shorter 75 bp fragments (10%) than longer 140 bp fragments (58%) (77). In their TB study, they reported 46% recovery of a 140 bp spiked synthetic TB target, but their extraction method is likely to fail for shorter TB cfDNA fragments expected in urine. Minimal recovery from TB-positive samples paired with what they hypothesize to be nonspecific amplification in TB-negative samples would explain their inability to detect TB-specific urine cfDNA.

The primary focus of my thesis project was to improve recovery of short urine cfDNA during the DNA extraction step by using sequence-specific purification, as discussed in Section 2.3.

2.2.3.3. PCR amplification targeting short fragments

The final requirement for detection of urine cfDNA is PCR designed to target short fragments. For successful amplification, a single, intact fragment must contain both PCR primer sites. PCR primers designed to amplify a shorter target will successfully amplify a larger fraction of fragments than PCR

primers spaced further apart (Figure 2.3C). At a minimum, the fragment length must be equal to the target amplicon length; however, this assumes that the primer sites are located precisely at the ends of the target fragment. In reality, cfDNA is randomly fragmented, and primer sites may fall anywhere along the length of the fragment. To ensure that as many fragments as possible are amplifiable, the PCR target length should be minimized. The clinical sensitivity of urine cfDNA detection increases with decreasing target length. Maximizing sensitivity by targeting shorter fragments is especially critical since urine cfDNA is also dilute, with total concentrations ranging from <1 ng/mL to 200 ng/mL (3, 47, 48) and copy numbers of specific targets much lower. In a study detecting *K-RAS* mutations in urine cfDNA, decreasing the PCR amplicon length from 157 bp to 87 bp increased clinical detection sensitivity from 20% to 67% (n=15) (78). This effect may be even more pronounced for fetal cfDNA, which is typically more fragmented than maternal cfDNA (45), and bacterial, viral, and mitochondrial cfDNA, which are not protected by histones and are therefore significantly more degraded than human genomic cfDNA (49, 53, 55). In a study detecting fetal cfDNA in maternal urine, decreasing PCR amplicon length from 65 bp to 39 bp increased clinical sensitivity from 25% to 75%. A further decrease to 25 bp was required before achieving 100% detection (n=10) (79). The sensitivity of Epstein-Barr virus detection similarly increased from 28% to 56% as the amplicon length decreased from 76 bp to 59 bp, and was accompanied by an increase in median detected cfDNA concentration from 290 copies/mmol creatinine to 7,040 copies/mmol creatinine (n=74) (80).

Most relevant to our application here, a modest 10 bp decrease in amplicon length (49 bp to 39 bp) for TB urine cfDNA led to more than 10-fold improvement in detected concentration (4 genome equivalents/mL to 50 genome equivalents/mL, n=1) (81). Minimizing PCR target length is critical to maximize detection sensitivity for TB urine cfDNA. Some previous TB urine cfDNA studies have targeted fragments as short as 38 bp (5, 6), but the majority used more conventional PCR designs targeting longer fragments (>150 bp) (17–19, 63–66, 68, 69, 71) or intermediate fragments (50–150 bp) (4, 59, 67, 70).

2.2.4. *Limitations of previous TB urine cfDNA studies*

Several review articles have identified and emphasized the impact of the methodological limitations of previous TB urine cfDNA detection (42, 52, 82), agreeing that low, variable sensitivity for urine cfDNA detection is in part due to sample preparation methods that discarded short fragments and/or amplification methods not designed to detect short fragments. Even for the limited number of studies more appropriately targeting short urine cfDNA, the methods used are generally not well described and their analytical performance (i.e., percent recovery, limit of detection) is not clearly characterized (4–6).

The lack of standardization in pre-analytical methodology, including the DNA extraction step, is a key limitation of urine cfDNA assays and has contributed to the lack of consensus on the value of urine cfDNA as a diagnostic target for TB.

In addition to methodological limitations, previous TB urine cfDNA studies have been, for the most part, limited in sample size and diversity in patient populations. One exception is Patel *et al.,* who directly compared sensitivity in populations with and without HIV co-infection with a total sample size of >400 (including controls). They observed slightly higher sensitivity in HIV-infected patients (45%, n=84) compared to HIV-uninfected patients (40%, n=90) (6). A critical question regarding urine cfDNA is whether it can be detected regardless of HIV status and CD4 count, unlike urine LAM. Past studies also included a majority of participants with smear-positive TB (95%, 60%, and 74% for Cannas *et al.,* Labugger *et al.,* and Patel *et al.,* respectively (4–6)), even though patients with smear-negative TB should be prioritized as a target population for whom non-sputum-based tests are urgently needed. Finally, urine cfDNA is a promising analyte for EPTB and pediatric TB, but these populations have not been included in many studies.

Another limitation is a lack of reliable quantification of the clinical range of TB urine cfDNA concentrations. Previous studies focused primarily on measuring diagnostic sensitivity and mostly neglected to report concentration ranges for TB urine cfDNA. Labugger *et al.* provided concentrations for a limited sample size, which ranged from 1–40.7 copies/ mL (median 6.8 copies/mL) for treatment naïve patients with detectable TB cfDNA (n=7) (5). For plasma cfDNA, Ushio *et al.* reported a median of 110 copies/mL (mean ± SD 7240 ± 26,916 copies/mL) (62). Without characterization of the sample preparation methods (i.e., percent recovery, variability across urine samples), however, the reported concentration ranges may not accurately reflect the actual urine cfDNA concentration.

2.2.5. *Potential for treatment monitoring*

Urine cfDNA has been suggested as a potential treatment monitoring tool for TB. Labugger *et al.* observed a spike in urine cfDNA levels within one week after treatment initiation and reduction in concentration over time to undetectable levels by week 12 in most patients (9/11) (5). They hypothesized that upon initiation of a successful treatment regimen, cfDNA concentration temporarily increases due to bactericidal activity, followed by a slow decline as the infection is cleared. More study of the dynamics of urine cfDNA over the course of active TB infection and treatment is needed.

2.2.6. *Potential for diagnosis of extrapulmonary TB*

Urine cfDNA may be particularly advantageous for diagnosis of EPTB, in which sputum is not always the appropriate or optimal sample type for testing. EPTB diagnosis typically requires a sample collected from the site of infection, which may be difficult to determine due to the varied clinical presentation of EPTB. The current evidence for EPTB diagnosis from urine cfDNA is limited but promising, suggesting possibly higher sensitivity than for pulmonary TB (PTB). Petrucci *et al.* reported a case study where TB cfDNA was detected in the urine of a pediatric patient with TB otitis media (83). They used the same methods described by Cannas *et al.* (no centrifugation, Wizard silica resin extraction, and moderate length 129 bp PCR target), which are moderately appropriate for short urine cfDNA (4). Torrea *et al.*, Rebollo *et al.*, da Cruz *et al.*, and da Costa Lima *et al.* tested the urine of both PTB and EPTB patients, albeit with non-optimal methods (i.e., discarded urine supernatant, long PCR amplicon). They all observed slightly higher detection sensitivities for EPTB compared to PTB (57.1% vs. 40.6%, 23% vs. 7%, 52% vs. 30%, and 40.6% vs. 35%, respectively) (66, 67, 69, 71). Fortún *et al.* also detected TB 16s RNA using the Gen Probe MTD assay in 70% (n=82) of EPTB patients but only 18% (n=28) of PTB patients (60). As expected, the sensitivity for genitourinary TB was high (89%), but the test was also able to detect lymphadentis TB (72%), miliary TB (67%), and multifocal TB (90%). The sample preparation methods were not well described.

Circulating TB cfDNA (in blood, not urine) has been detected in EPTB patients using digital PCR (dPCR). Yamamoto *et al.* described a case study where IS6110 TB cfDNA (65.5 copies/mL) was detected in the plasma of a patient with disseminated TB after a hematopoietic stem cell transplantation (84). Culture, smear, and NAAT testing of sputum, blood, and urine samples were all negative, confirming the cell-free nature of the detected TB DNA and suggesting that cfDNA testing may offer diagnostic value for EPTB when other methods fail. Yang *et al.* compared detection of circulating TB cfDNA from patients with PTB and EPTB (including tubercular meningitis, abdominal TB, tuberculosis pleurisy, tuberculosis lymphadenitis of the neck, bone TB, and pelvic TB) (85). The concentration of detected DNA was similar for PTB and EPTB (201.80 ± 40.94 copies/µL *vs.* 167.40 ± 40.82 copies/µL). There was no difference in copy number between different types of EPTB. This study analyzed whole blood and used phenol-chloroform extraction, so it is not guaranteed that exclusively cfDNA was detected. The reported concentrations are noticeably higher than expected for circulating cfDNA, although it is not clear whether they are given per microliter of blood, purified DNA, or dPCR reaction. Both studies used amplicons short enough to detect plasma cfDNA (71 bp and 83 bp, respectively), which is less degraded

than urine cfDNA. Although these studies did not test for urine cfDNA, a fraction of the detected circulating cfDNA would be expected to be excreted in the urine.

2.2.7. *Potential for diagnosis of pediatric TB*

Although children are a key underserved population who especially stand to benefit from a sensitive non-sputum-based TB test, the study of TB urine cfDNA in pediatric patients has been limited. As mentioned above, a single case study detected TB urine cfDNA in a pediatric patient with TB otitis media who had negative smear microscopy and Xpert MTB/RIF from sputum, gastric aspirates, and nasopharyngeal aspirates (83). Another cohort study detected TB DNA in the urine of pediatric patients with 39% sensitivity for pulmonary TB and 20% sensitivity for extrapulmonary TB, but did not use methods capable of detecting short fragments (86). Specifically, they used a nested PCR design (409/316 bp) targeting longer fragments, so it is more likely that they detected cell-associated DNA rather than cfDNA.

2.2.8. *Ongoing optimization of pre-analytical variables for TB urine cfDNA*

The optimal storage conditions to preserve urine cfDNA are unclear, with mixed results from a limited number of studies. One study measuring the degradation of human genomic cfDNA during storage reported that stability was dependent on the patient population and/or collection site (87). For one population (in Italy), addition of 40 mM EDTA prevented degradation (mean 1.6% loss) for 28 days storage at room temperature, 4°C, and -80°C, but not -20°C (94.7% loss). For another population (in Zambia), addition of 10 mM EDTA had no effect on stability regardless of storage temperature (99.3% loss after 28 days, the majority of which was during the first 7 days). Another study found that addition of preservatives was critical for preventing degradation of pathogen cfDNA, including TB cfDNA (88). They found that 25 mM EDTA was superior to Streck urine preservative for a processing delay of up to 24 hours at room temperature. For long term storage (24 weeks at –80°C), choice of preservative had no effect. A third study found that the effect of EDTA was superior to that of cold storage for preserving urine cfDNA (89). Over the course of 72 days storage, cfDNA concentration declined for all conditions tested (no EDTA vs. 40 mM EDTA, room temperature vs. -20°C) but least so for urine stored with 40 mM EDTA at -20°C. Yet another study found that 40 mM EDTA reduced degradation at room temperature, but that cfDNA was preserved regardless of whether or not EDTA was added when stored at 4°C, -20°C, or -80°C for 28 days (90). Yet another study found that samples stored for 3 months at −70°C with 10 mM EDTA resulted in the highest cfDNA (compared to no EDTA or -20°C storage) (91). A final study found that Streck urine preservative performed best for 1 week storage at room temperature, and that

4°C storage was superior to room temperature (92). EDTA was not included in the comparison. Collectively, these studies suggest that cold storage with the addition of EDTA is ideal for urine cfDNA preservation, but cold storage alone may be sufficient. The required EDTA concentration and differences (if any) between 4°C, -20°C, and -80°C storage remain unclear. Variability in results suggest likely heterogeneity in cfDNA stability across urine samples.

A few studies have compared clinical detection rates (79, 81) and total cfDNA recovery (48, 91, 93) using a limited set of cfDNA extraction methods. For lab-based protocols, a study determined that Q Sepharose resin improves extraction of short cfDNA compared to Wizard silica resin (79, 81). For commercially-available kits, one study identified the Qiagen QIAamp Viral RNA Mini Kit as the best performing kit overall (93), another identified the Applied Biosystems MagMAX Cell-Free DNA Isolation and Norgen Urine Cell-Free Circulating DNA Purification Midi kits as having the highest yields of 50–100 bp cfDNA (91), and a third identified the Perkin Elmer NEXTprep-Mag Urine cfDNA Isolation Kit as similar but slightly superior to the Norgen Urine Cell-Free Circulating DNA Purification Midi Kit (48).

Two studies compared additional pre-analytical variables, including timing of urine collection and centrifugation protocol for isolation of cfDNA from cell-associated sediment (88, 92). The most notable finding was that the first 50 mL urine void contained, on average, the most cfDNA (92). Continued optimization of pre-analytical variables is needed to ensure best results from urine cfDNA analysis. Development of a well-characterized, quantifiable assay will contribute to the continued investigation of these unanswered questions.

2.3. Sequence-specific purification

This section is adapted from

Oreskovic A, Lutz BR. Ultrasensitive hybridization capture: reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. PLOS ONE 16(2): e0247851 (2021). https://doi.org/10.1371/journal.pone.0247851.

We identified sequence-specific purification (interchangeably referred to here as hybridization capture) as a method likely to perform well for dilute, fragmented urine cfDNA. Sequence-specific purification uses oligonucleotide probes that are complementary to the sequence(s) of interest to capture target-specific nucleic acids via hybridization. The capture probes are immobilized on a substrate (commonly magnetic beads) to allow for isolation, washing, and concentration of target-specific nucleic acid. Generally, there are two hybridization capture formats distinguished by the order of the target-probe

hybridization and probe immobilization steps. The first, referred to here as solution-based hybridization, involves hybridization of biotinylated probes to target in solution, followed by capture of target-probe duplexes on streptavidin-coated magnetic beads (Figure 2.5A). The second, referred to here as surface-based hybridization, pre-immobilizes probes on magnetic beads via biotin-streptavidin binding or covalent coupling. The probe-bead complexes are then added to the sample, where target hybridizes directly to the immobilized probes (Figure 2.5B).

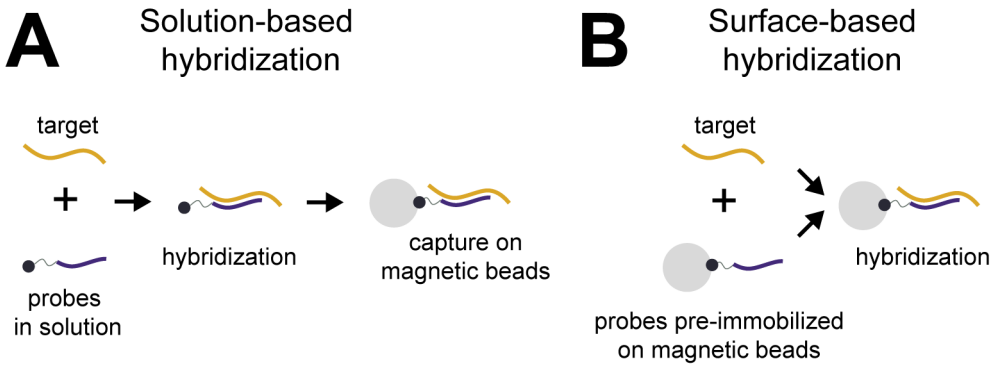


Figure 2.5: Solution-based versus surface-based hybridization capture. (A) *Solution-based hybridization hybridizes probes to target in solution, followed by capture of probe-target complexes on magnetic beads.* ***(B)*** *Surface-based hybridization hybridizes target to probes that are pre-immobilized on magnetic beads.*

2.3.1. *Previous uses of sequence-specific purification for sample preparation*

Sequence-specific purification has been previously used for capture of nucleic acids directly from a variety of sample types, as summarized in Table 2.2. In these studies, sequence-specific purification was used as the primary sample preparation method to enrich target nucleic acid directly from raw samples or crude lysates with minimal processing (such as lysis and/or centrifugation, but no precipitation or column purification). We excluded studies that enriched bacterial samples by culture (94–97), pre-purified nucleic acids (98–104), or pre-amplified nucleic acids before hybridization (e.g., targeted enrichment of sequencing libraries (105)). From this survey of previous applications of sequence-specific purification for sample preparation, the versatility and potential high sensitivity and specificity of the technique are clear. Sequence-specific purification is tolerant to a wide variety of clinical sample matrices and even performs well in the presence of lysis buffers. It may be especially advantageous for samples with high concentrations of amplification inhibitors (including but not limited to non-target genomic DNA) or targets that require harsh denaturants for lysis. Addition of chaotropic salt such as guanidinium thiocyanate can simultaneously lyse cells and viruses and preserve fragile nucleic acids, enabling targeting of RNA as well as DNA (106, 107). While manual, magnetic bead protocols have been

most commonly used, sequence-specific purification has also been carried out in microfluidic (108) and lateral flow (109) formats.

The best reported analytical sensitivities are in the range of 5–100 copies, genome equivalents, or colony forming units, with a maximum sample input of 1 mL. Purification efficiencies, when reported, were variable (<10% to 97%) but suggest the potential for high recovery if conditions are optimized. Unfortunately, many previous studies did not report detailed protocols or characterization of their methods (limit of detection, percent recovery, etc.), making direct comparison difficult. Non-optimal probe concentrations, bead volumes, and probe:bead ratios may have resulted in differences in performance across protocols. Reports on the performance of solution-based versus surface-based hybridization are inconsistent, with no consensus on when one approach may be preferred to the other.

Table 2.2: Previous uses of sequence-specific hybridization capture for sample preparation without bacterial culture, pre-purification, or pre-amplification.

Study	Hybridization format[a]	Target	Sample type(s)	Sample input volume[b]	Percent recovery	Analytical sensitivity
Albretsen *et al.* (1990) (106)	Surface	Measles morbillivirus RNA	Crude cell lysate	Not reported	40 – 45%	630 amol (3.8×10^8 copies)
Muir *et al.* (1993) (110)	Surface	Enterovirus RNA	CSF, stool, saliva, blood, pericardial fluid, urine, solid tissue	100 µL	Not reported	Not reported
Heermann *et al.* (1994) (111)	Solution	Hepatitis B DNA	Serum	50 µL	Not reported	10 – 100 GE/mL
Jacobsen (1995) (112)	Surface	*Psuedomonas fluorescens* DNA	Rhizosphere soil	0.5 mg	Not reported	50 CFU
Mangiapan *et al.* (1996) (113)	Solution	*Mycobacterium tuberculosis* DNA	Pleural fluid	550 µL	Not reported	5 – 10 GE
Beaulieux *et al.* (1997) (107)	Surface	Enterovirus RNA	Virus supernatant	100 µL	Not reported	300 GE/mL
Brugiere *et al.* (1997) (114)	Solution	*Mycobacterium tuberculosis* DNA	Bronchoalveolar lavage fluid	550 µL	Not reported	Not reported
Arnal *et al.* (1999) (115)	Surface	Hepatitis A RNA	Stool and shellfish	200 µL	Not reported	Not reported
Amagliani *et al.* (2006) (116)	Surface	*Listeria monocytogenes* DNA	Milk	Not reported	Not reported	10 CFU/mL
Thompson *et al.* (2006) (117)	Solution	*Salmonella* DNA	Water	20 µL	47%	5 cells
Yeung *et al.* (2006) (108)	Solution and surface (separately; microfluidic device)	*Escherichia coli* DNA	Cell lysates	2 µL	75% (solution); 30% (surface)	100 – 1000 cells/mL (solution)
Parham *et al.* (2007) (118)	Surface	Group B Streptococci DNA	Vaginal and anal swabs	<30 µL	<10% (≥10,000 copies) – 60% (100 copies)	1250 CFU/mL
Vansnick *et al.* (2007) (119)	Solution	*Mycobacterium avium* subspecies *paratuberculosis* DNA	Stool and tissue	500 µL	Not reported	100 bacilli

Peeters *et al.* (2012) (120)	Solution and surface (separately)	Human papillomavirus DNA	Vaginal swabs	Not specified (<200 µL)	4% (solution), 25% (surface)	Not reported
Rodriguez *et al.* (2012) (121)	Solution	*Batrachochytrium dendrobatidis* DNA	Amphibian skin swabs	10 µL	Not reported	Not reported
Adams *et al.* (2015) (122)	Surface	Respiratory syncytial virus RNA	N/A	25 µL	20 (30 min) – 77% (3 hr)	Not reported
Rohrman *et al.* (2015) (109)	Surface (lateral flow)	Human immunodeficiency virus DNA	Blood	20 µL	Not reported	10,000 copies
Guo *et al.* (2015) (123)	Surface	Hepatitis B DNA	Serum	10 µL	74% (100 copies) – 97% (10^6 copies)	90 IU/mL
Reed *et al.* (2016) (124)	Solution	*Mycobacterium tuberculosis* DNA	Sputum	950 µL	Not reported	20 CFU/mL (340 copies/mL)
Reed *et al.* (2017) (125)	Solution	*Mycobacterium tuberculosis* DNA	Sputum	950 µL	Not reported	5 CFU/mL (85 copies/mL)

GE = genome equivalents; CFU = colony forming units; IU = international unit

[a] *All studies used magnetic beads unless otherwise noted.*

[b] *Sample input volume is given as volume after any concentration or centrifugation steps, if relevant.*

2.3.2. *Previous uses of sequence-specific purification for TB diagnosis*

Sequence-specific purification has previously been used to improve detection of TB DNA in paucibacillary respiratory samples, first by Mangiapan *et al.* and Brugière *et al.* in 1996–1997. Mangiapan *et al.* detected TB DNA purified by sequence-specific purification targeting IS6110 and the direct repeat (DR) region in the pleural fluid of 13 of 17 patients with tuberculosis pleurisy, including 3/3 culture-positive samples and 10/14 culture-negative samples (113). When DNA was purified by silica adsorption instead, only the 3 culture-positive samples were detectable. The authors were able to reliably detect 5–10 genome equivalents, and occasionally detect down to 1 genome equivalent, of TB DNA in samples with up to 750 µg total DNA, a 10- to 100-fold increase in sensitivity compared to amplification of total DNA purified by silica adsorption. Brugière *et al.* detected TB DNA in all bronchoalveolar lavage samples from smear-negative pulmonary TB patients (n=9) (114). More recently, Reed *et al.* have published two consecutive studies using sequence-specific purification to improve detection of TB DNA in sputum samples. In the first study, they identified human genomic DNA as a significant qPCR inhibitor in DNA purified from sputum samples (124). The inhibitory effect was eliminated by DNase treatment of samples prior to spiking in target DNA. To overcome this problem, they developed a hybridization capture method targeting IS6110 and *senX3-regX3*, that, combined with a new sputum thinning procedure, had an estimated limit of detection of 20 CFU/mL from 1 mL sputum. The H37Ra strain they used had 17 copies of IS6110 per genome, so the limit of detection corresponds to 340 copies/mL of IS6110. In a clinical pilot study, the sensitivity and specificity of the assay were 96% and 100%, respectively, compared to Xpert MTB/RIF (n=60). In the subsequent study, they improved upon the sequence-specific purification method and reported the results of an expanded clinical study

(125). The limit of detection of their test was 5 CFU/mL sputum (corresponding to 85 copies/mL IS6110), approaching that of culture (10–100 CFU/mL). The clinical sensitivity and specificity were 94.9% and 100%, respectively, compared to culture (n=142). The test had slightly lower sensitivity in smear-negative (88.6%) compared to smear-positive (98.4%) samples.

2.3.3. *Rationale for application of sequence-specific purification to urine cfDNA*

In previous implementations, sequence-specific purification was used primarily to isolate target DNA from an excess of non-target DNA, which can inhibit downstream amplification and reduce sensitivity, particularly for dilute targets in an excess of non-target DNA (109, 113, 124). For this reason, sequence-specific purification has been commonly used for environmental samples and pathogen detection in samples with large amounts of human gDNA (e.g., stool, sputum, whole blood). For urine cfDNA, however, the total nucleic acid concentration is low and unlikely to inhibit downstream amplification. In this case, we instead aimed to leverage hybridization's ability to sensitively capture and retain short fragments regardless of length and concentration. We hypothesized that sequence-specific purification would increase recovery of short, dilute cfDNA relative to silica-based extraction methods, which would in turn increase the diagnostic sensitivity for detection of TB urine cfDNA. Removal of non-target DNA (for example, cell-associated human gDNA in urine samples that are not clean catch) may be a secondary benefit of sequence-specific purification by helping to minimize the likelihood of nonspecific amplification. To our knowledge, sequence-specific purification has not been employed previously to target urine cfDNA.

2.4. GAPS IN THE LITERATURE AND PROJECT MOTIVATION

There is clearly a need for rapid TB diagnostics that do not rely on sputum samples. Urine cfDNA is an attractive target for TB diagnosis, but research regarding its implementation remains in the early stages. More clinical evidence, molecular characterization, and the determination of optimal methodology are needed to establish urine cfDNA as a viable diagnostic target for TB. There are several areas where this dissertation seeks to contribute both broadly to the field of urine cfDNA, and more specifically to improve TB diagnosis from urine cfDNA:

- Determination of optimal sample preparation methods for urine cfDNA
- Development of a sequence-specific purification method suitable for TB urine cfDNA by improving upon sensitivity and sample volume of previous sequence-specific purification methods

- Design of PCR assays for specific amplification of short cfDNA
- Assessment of the impact of improved methods on sensitivity of TB diagnosis from urine cfDNA
- Application of a sensitive, well-characterized, and quantifiable method to study TB urine cfDNA in new clinical populations
- Characterization of the concentration, composition, and fragment length distribution of TB urine cfDNA to inform continued development of diagnostic assays

3. Development of sequence-specific purification method for urine cell-free DNA

Portions of this chapter are adapted from

Oreskovic A, Lutz BR. Ultrasensitive hybridization capture: reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. PLOS ONE 16(2): e0247851 (2021). https://doi.org/10.1371/journal.pone.0247851.

This paper was written to target a broader audience interested in sequence-specific purification of cfDNA for applications beyond TB. Here, the text has been modified and sections have been added to focus specifically on the development of an assay to diagnose TB from urine cfDNA.

3.1. Abstract

Urine cell-free DNA (cfDNA) is a promising biomarker for non-sputum-based tuberculosis (TB) diagnosis, but there is no consensus on its optimal pre-analytical methodology, including the DNA extraction step. Due to its short length (majority of fragments <100 bp) and low concentration (ng/mL), urine cfDNA is not efficiently recovered by conventional silica-based extraction methods. To maximize sensitivity of TB urine cfDNA detection, we developed an ultrasensitive hybridization method that uses sequence-specific oligonucleotide capture probes immobilized on magnetic beads to improve extraction of short cfDNA from large-volume urine samples. Our sequence-specific purification method recovers near 100% (95% CI: 82.5–105.8%) of TB target-specific DNA from 10 mL urine, independent of fragment length (25–150 bp), and has a limit of detection of ≤5 copies of double-stranded DNA (0.5 copies/mL). Pairing sequence-specific purification with an ultrashort qPCR design, we can efficiently capture and amplify fragments as short as 25 bp. Our method enables amplification of TB cfDNA from 10 mL urine in a single qPCR well, tolerates variation in sample composition, and effectively removes non-target DNA. Our sequence-

specific purification protocol improves upon both existing silica-based urine cfDNA extraction methods and previous hybridization-based sample preparation protocols. Two key innovations contribute to the strong performance of our method: a two-probe system enabling recovery of both strands of double-stranded DNA and dual biotinylated capture probes, which ensure consistent, high recovery by facilitating optimal probe density on the bead surface, improving thermostability of the probe-bead linkage, and eliminating interference by endogenous biotin. While we developed the sequence-specific purification method specifically to improve TB diagnosis from urine cfDNA, we expect that it will be versatile across urine cfDNA targets and may be useful for other cfDNA sample types and applications beyond cfDNA. To make our sequence-specific purification method accessible to new users, we present a detailed protocol and straightforward guidelines for designing new capture probes.

3.2. Introduction

Urine cell-free DNA (cfDNA) is generated from transrenal excretion of circulating cfDNA, as well as from local degradation of cells shed along the urogenital tract (3). It is a versatile noninvasive biomarker with applications in cancer detection (3, 38, 126), infectious disease diagnosis (4, 42), organ transplant monitoring (43, 44), and prenatal genetic screening (3, 45). To ensure reliable outcomes of urine cfDNA assays, optimal pre-analytical methodology is essential. Several studies have compared collection procedures, storage conditions, and sample preparation methods for urine cfDNA (48, 79, 88, 91, 92), but there remains no consensus on best practices, and the conclusions for more well-studied plasma cfDNA may not apply to urine cfDNA. Here, we aimed to address deficiencies in one of many critical pre-analytical variables that may impact the performance of urine cfDNA tests: the DNA extraction step.

Conventional silica-based DNA extraction methods are not suitable for urine cfDNA, which is dilute and extensively fragmented. While plasma cfDNA has a peak fragment length of approximately 167 bp, reflecting protection of histone-associated DNA within nucleosomes (51), urine cfDNA is more fragmented due to glomerular filtration of the transrenal fraction and fast degradation kinetics of all cfDNA in urine. The distribution of fragment lengths varies across samples, but the majority of urine cfDNA fragments are expected to be <100 bp (45, 54–56), with a wider distribution of fragments around the peak fragment length compared to plasma cfDNA (56). Some forms of urine cfDNA, including fetal (45), tumor (57), mitochondrial (45, 55), microbial (55), and viral (55) cfDNA, are even more fragmented than human nuclear cfDNA. For these particularly short forms of urine cfDNA, the peak fragment length may be as short as 30 to 60 bp (45, 55). Both DNA extraction and sequencing library preparation methods may underestimate the prevalence of short fragments, making it difficult to determine the true

fragment length distribution of urine cfDNA. Recently, new single-stranded library preparation methods for next-generation sequencing have revealed that very short, formerly undetectable fragments make up a larger fraction of both plasma (49, 51, 58) and urine (55) cfDNA than previously recognized. Although thorough characterization of the full diversity of urine cfDNA using extraction and library preparation methods sensitive to the shortest fragments remains incomplete, it is increasingly clear that targeting short fragments is critical to maximizing the sensitivity of urine cfDNA assays.

As further evidence of the importance of targeting short fragments, decreasing PCR amplicon length has been shown to increase both clinical sensitivity and the detected concentration of urine cfDNA (79–81). Detection sensitivity for fetal cfDNA in maternal urine was 25%, 75%, and 100% for PCR amplicon lengths of 65 bp, 39 bp, and 25 bp, respectively (79). The sensitivity of Epstein-Barr virus detection similarly increased from 28% to 56% as the amplicon length decreased from 76 bp to 59 bp, and was accompanied by a 24-fold increase in median detected urine cfDNA concentration (80). In another study, a modest 10 bp decrease in PCR amplicon length from 49 bp to 39 bp resulted in more than 10-fold increase in the detected concentration of tuberculosis (TB) cfDNA in urine (81).

The ability to detect short fragments is particularly important given the low total concentration of cfDNA in urine (<1–200 ng/mL) (3, 47, 48) and potential for very low copy numbers of target-specific cfDNA. To ensure maximum sensitivity for detection of low-concentration, fragmented cfDNA, it is critical not only to amplify short targets, but also to use a DNA extraction method capable of retaining short fragments with high efficiency. Conventional DNA extraction methods based on adsorption of DNA to silica under chaotropic conditions (73) are inadequate for urine cfDNA because recovery decreases with decreasing fragment length and concentration. The driving forces of DNA adsorption to silica – namely hydrophobic interactions as a result of dehydration of silica and DNA surfaces and hydrogen bonding between silica and the DNA backbone – are proportional to DNA fragment length (74). The lower length limit of silica-based methods varies across binding matrices and buffer conditions, but recovery generally decreases for fragments less than 50–100 bp. Silica-based methods may also perform poorly for dilute samples like urine cfDNA because a non-trivial fraction of DNA may remain irretrievably bound to the silica surface (75, 76). We previously compared several silica-based methods for urine cfDNA extraction, both commercially-available kits and lab-based protocols, and discovered that none maintained high recovery across all fragment lengths from 25–150 bp (see Chapter 4) (127). Many had low or undetectable recovery of 25–40 bp fragments (127), which are expected to make up a significant fraction of urine

cfDNA. Even if all other pre-analytical variables are optimized, low recovery of short fragments during the DNA extraction step risks compromising clinical results.

We aimed to develop an improved method for extraction of short cfDNA from urine. Our goal was to improve purification and detection of known target sequences for diagnostic applications, specifically TB, rather than to extract total urine cfDNA. We identified sequence-specific hybridization capture as a technique likely to perform well for short, dilute urine cfDNA because it should be agnostic to DNA fragment length and concentration.

A sequence-specific purification method for short, dilute urine cfDNA must meet two design criteria not achieved with previous sequence-specific purification methods. Due to the very low concentration of TB cfDNA in urine (median 15 copies/mL in our clinical study), I needed to push the limits of sequence-specific purification to 1) achieve <1 copy/mL copy sensitivity and 2) increase the input sample volume to 10 mL. In contrast, previous sequence-specific purification methods (Table 2.2) had, at best, limits of detection of 5–100 copies/mL and were limited to a sample volume of ≤1 mL.

Here, we report the development and optimization of an ultrasensitive hybridization method that uses dual biotinylated sequence-specific probes immobilized on streptavidin-coated magnetic beads to capture, concentrate, and purify target-specific cfDNA fragments from 10 mL urine samples. Our sequence-specific purification method is intended for sample preparation prior to qPCR-based detection of known target sequences, such as in the case of pathogen detection; it is not suitable for applications requiring purification of total cfDNA prior to sequencing. We have identified key parameters affecting hybridization efficiency and implemented designs to ensure consistent, high percent recovery. We report full characterization of the analytical performance of our sequence-specific purification method and describe two key innovations that contribute to its unprecedented sensitivity: dual biotinylated capture probes and a two-probe system for recovery of both strands of double-stranded DNA (dsDNA). Our sequence-specific purification method improves upon the analytical performance of both existing silica-based urine cfDNA extraction methods and sequence-specific purification protocols, including our own previously-reported 1 mL solution-based hybridization assay (127). Although we developed the sequence-specific purification method for TB diagnosis and focused our design criteria on the needs of urine cfDNA (i.e., short, dilute fragments), we anticipate that our method will be versatile across urine cfDNA applications and may also be useful for other sample types. With the aim of making our method readily accessible to new users and applicable to broader potential applications of sequence-specific

purification, we give a detailed, optimized protocol and outline guidelines for straightforward design of capture probes for new targets.

3.3. Development of sequence-specific assay for TB urine cfDNA

3.3.1. *Selection of genomic target*

Our TB urine cfDNA assay targets the *Mycobacterium tuberculosis* (MTB) insertion sequence IS6110 (GenBank Accession #X17348), which is a well-studied target for TB diagnosis (128, 129). It is present at variable copy number (1–28 inserts) in ~99% of MTB complex strains (Figure 3.1, unpublished data courtesy of Dr. Norman D. Brault), providing an opportunity for enhanced sensitivity compared to single-copy genomic elements. I designed capture probes and PCR primers to target a 40 bp amplicon within a conserved and specific region of IS6110 (130).

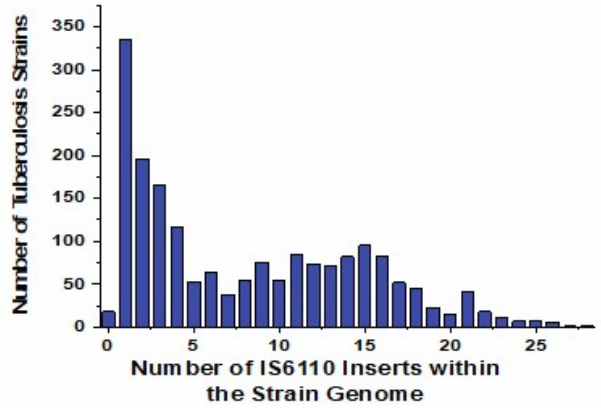


Figure 3.1: Copy number of IS6110 genomic target across TB strains. *Genomes of ~1900 MTB complex strains obtained from NCBI GenBank and analyzed for IS6110 presence and prevalence. IS6110 was present in approximately 99% of MTB complex strains. Data courtesy of Dr. Norman D. Brault.*

3.3.2. *Overview of sequence-specific purification method and probe design*

The hybridization capture probe design is shown in Figure 3.2A**.** Two unique capture probes are used to hybridize to both strands of the dsDNA target region. Each probe contains a short sequence-specific binding region (15–20 bp) complementary to the target of interest and a 5’ dual biotin modification for immobilization on streptavidin magnetic beads. The probes also contain a 20 bp poly(A) spacer between the dual biotin modification and target-specific binding sequence to reduce steric hindrance during hybridization and a 3’ C3 blocker to prevent extension of any residual probes during qPCR. As discussed in later sections, both the dual biotin modification of probes and the two-probe system for recovery of both strands of dsDNA are critical design elements for the strong analytical performance of our method.

An overview of the sequence-specific purification procedure is given in Figure 3.2B. Dual biotinylated capture probes complementary to the target of interest are immobilized on streptavidin-coated magnetic beads prior to hybridization. The functionalized beads are added to a 10 mL urine sample, along with NaCl (1M) to thermodynamically encourage target-probe hybridization and Tween-20 (0.1%) to improve probe accessibility near the bead surface. A relatively high effective probe concentration (2 nM of each probe) helps drive rapid hybridization. Urine is denatured at >90°C for 15 minutes before the DNA target sequence is captured via hybridization to probes at room temperature for 30 minutes. The beads are concentrated by centrifugation, then washed on a magnetic rack, twice with a high salt buffer and once with a low salt stringency buffer, to remove urine and non-target DNA. The purified target-specific DNA is eluted under basic conditions, neutralized, and detected by short-target qPCR. The entire purified output (20 µL) from 10 mL urine can be analyzed in a single qPCR well, resulting in 500X concentration from the original sample volume and minimizing dilution to enable detection of low concentration targets.

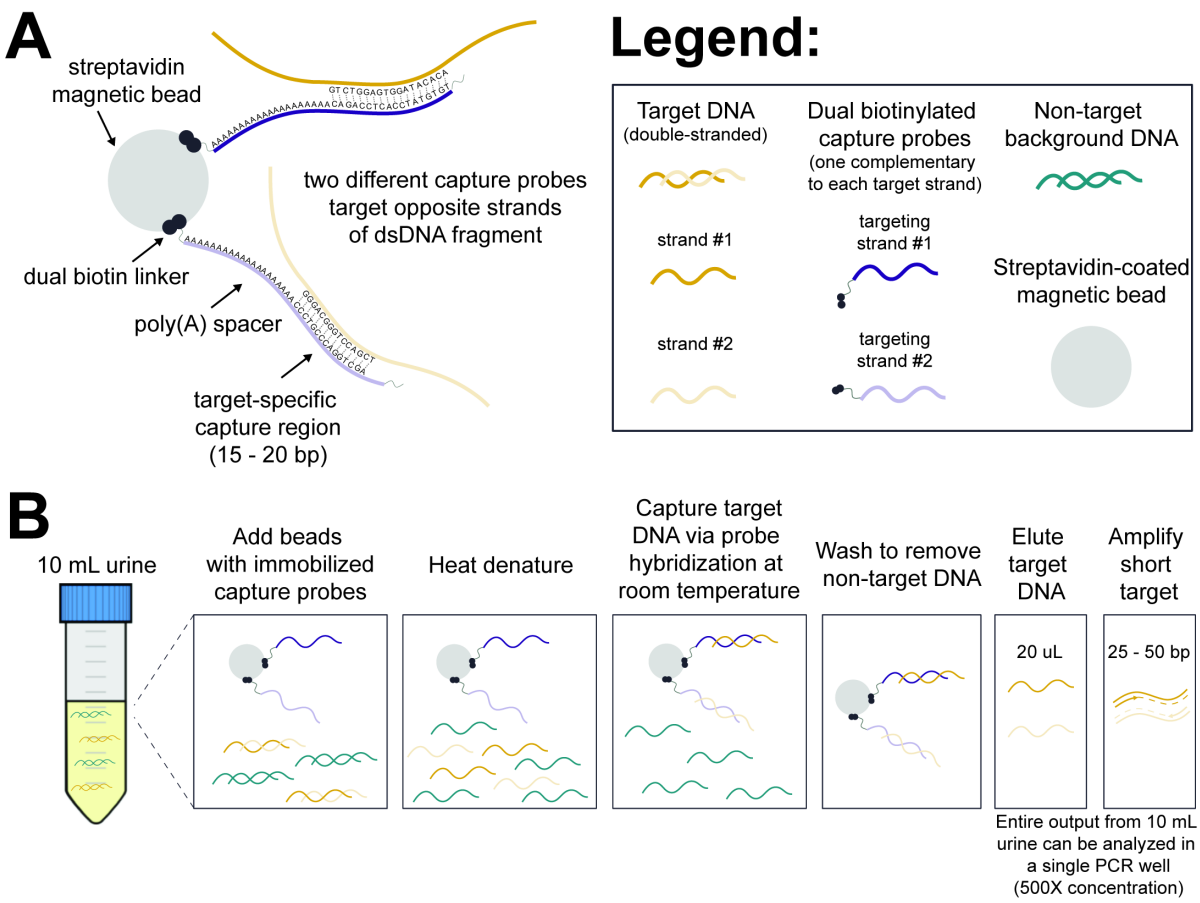


Figure 3.2: ***Hybridization capture probe design and sequence-specific purification procedure. (A)*** *Schematic illustrating capture probe design. Two short, sequence-specific probes, one complementary to*

each strand of the double-stranded target region, are immobilized on streptavidin-coated magnetic beads via dual biotin linkers. The target-specific capture regions are based on truncated versions of the PCR primers. In addition to the 5' dual biotin modification, each probe has a 20 bp poly(A) spacer to reduce steric hindrance near the bead surface and a 3' C3 spacer to prevent extension. ***(B)*** *Overview of hybridization capture procedure. Beads with pre-immobilized capture probes are added to 10 mL urine (along with 1 M NaCl and 0.1% Tween-20), heat denatured (15 minutes at >90°C), hybridized to target DNA (30 minutes at room temperature), and washed to remove non-target DNA (2X high-salt wash, 1X low-salt wash). Purified target-specific DNA is eluted under basic conditions, neutralized, and amplified by short-target qPCR.*

3.4. SEQUENCE-SPECIFIC PURIFICATION PROTOCOL

A full step-by-step sequence-specific purification protocol is posted on protocols.io (http://dx.doi.org/10.17504/protocols.io.bep4jdqw). The protocol below was used for analytical characterization of the hybridization capture method, with experimental details and any modifications for individual experiments noted in the following Methods section.

I. **Collect, store, and prepare urine cfDNA**

1. Before urine collection, prepare 15 mL DNA LoBind tubes (Eppendorf, Hamburg, Germany) with 500 µL 0.5 M EDTA and 100 µL 1M Tris-HCl pH 8.
2. Collect urine sample in a sterile container.
3. Immediately after urine collection, add 10 mL of urine to each prepared 15 mL tube and mix by inversion. The final concentration will be 25 mM EDTA and 10 mM Tris-HCl.
4. Freeze at -80°C until analysis.
5. Immediately before analysis, thaw urine at 37°C and mix by inversion.
6. Optionally, test with Fisherbrand 10-SG Urine Reagent Strips (Thermo Fisher Scientific, Waltham, MA, USA).
7. Centrifuge urine for 5 minutes at 8000g to pellet cell debris.
8. Transfer cell-free urine supernatant to new 15 mL DNA LoBind tube (Eppendorf).

II. **Immobilize capture probes on magnetic beads**

1. Vortex Dynabeads MyOne Streptavidin C1 (Thermo Fisher) for 30 seconds to ensure that beads are evenly dispersed in solution.
2. Pipette beads into a 1.5 mL DNA LoBind tube (Eppendorf). Prepare 50 µL (0.5 mg) beads per 10 mL urine sample to be analyzed.
3. Wash beads three times with an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
    i. Place beads on magnetic rack for 1 minute; remove and discard supernatant.

ii. Add an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20), vortex for 5 seconds, and briefly spin down.

iii. Repeat twice for a total of three washes.

4. Resuspend beads in an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
5. Pre-mix dual biotinylated capture probes BP1 and BP2 (Table 3.1) in TLE to a final concentration of 50 µM BP1 and 50 µM BP2.
6. Add pre-mixed dual biotinylated capture probes BP1 and BP2 to beads and immediately vortex for 5 seconds. Use 25 pmol of each probe per 50 µL beads (0.5 µL of pre-mixed stock with 50 µM of each probe). If using only a single probe, use 50 pmol per 50 µL beads.
7. Rotate for 15 minutes at room temperature to immobilize probes on beads.
8. Briefly spin down.
9. Wash beads three times with an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
    i. Place beads on magnetic rack for 1 minute; remove and discard supernatant.
    ii. Add an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20), vortex for 5 seconds, and briefly spin down.
    iii. Repeat twice for a total of three washes.
10. Resuspend beads in an equal volume high salt wash buffer.

III. **Capture TB-specific urine cfDNA by hybridization**

1. Add 2.5 mL 5 M NaCl (final concentration 1 M), 127 µL 10% Tween-20 (final concentration 0.1%) and 50 µL prepared beads to each 10 mL urine sample. If spiking in positive control (Table 2), add it now.
2. Mix well by inversion.
3. Denature for 15 minutes in dry bath with 15 mL tube block preheated to 120°C (urine temperature should reach >90°C). Alternatively, a 95°C water bath can be used.
4. Rotate for 30 minutes at room temperature to hybridize target cfDNA to capture probes.

IV. **Wash to remove urine inhibitors and non-target DNA**

1. Centrifuge for 5 minutes at 5000g to pellet beads.
2. Remove and discard all but ~1 mL supernatant using 10 mL serological pipette.

3. Resuspend beads in remaining supernatant and transfer to 1.5 mL DNA LoBind tube (Eppendorf).
4. Place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.

    Note: Infrequently, clinical urine samples with blood and/or proteinuria may cause beads to aggregate poorly on the magnet. If this is the case, using a minicentrifuge to help pellet the beads before removing the supernatant or leaving the tube on the magnet for >1 minute here and in subsequent wash steps may be helpful. It is okay to leave more residual volume to avoid bead loss (see Section 7.6 for more information and suggested addition of proteinase K digestion step).
5. Add 1 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and wash by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
6. Place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.
7. Add 1 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and wash by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
8. Place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.
9. Add 1 mL low salt wash buffer (15 mM NaCl, 10 mM Tris-HCl pH 8, no Tween-20) and wash by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
10. Place on magnetic rack for 1 minute. Remove and discard supernatant.
11. Spin down again, place on magnetic rack, and remove as much liquid as possible using P20 pipette.

V. **Elute purified TB cfDNA**

1. Add 20 µL freshly prepared 20 mM NaOH, vortex for 5 seconds, and spin down briefly.
2. Place on magnetic rack. Transfer as much supernatant as possible (usually 20–21 µL) directly to qPCR well or to new DNA LoBind tube (Eppendorf). This contains purified target cfDNA. Avoid transferring any beads to qPCR.

Note: Infrequently, clinical urine samples with blood and/or proteinuria may make it difficult to avoid bead transfer to PCR. If this is the case, using a minicentrifuge to help pellet the beads before eluting or leaving the tube on the magnet for longer may be helpful. It is okay if the eluted volume is <20 µL or if a small amount of beads are transferred to PCR, although this will lead to PCR inhibition (see Section 7.6 for more information and suggested addition of proteinase K digestion step).

3. Partially neutralize with 3.5 µL 100 mM HCl. The qPCR buffer will adjust to final pH and tolerates slightly basic pH better than acidic pH.
4. Proceed directly to qPCR.

VI. **Quantify by qPCR**

1. Analyze the entire hybridization output (approximately 24 µL) from each 10 mL urine sample in a single qPCR well. Each 50 µL reaction should contain 1.25 U OneTaq Hot Start DNA Polymerase (New England Biolabs [NEB], Ipswitch, MA, USA), 1X NEB OneTaq GC Reaction Buffer (NEB; 80 mM Tris-SO_4, 20 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 5% glycerol, 5% DMSO, 0.06% IGEPAL CA-630, 0.05% Tween-20, pH 9.2), 0.8 mM dNTPs (NEB), 0.4X EvaGreen (Biotium, Fremont, CA, USA), 200 nM forward primer (Table 3.1), and 200 nM reverse primer (Table 3.1).
2. Amplify in CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) using an initial denaturation of 94°C for 3 min followed by 45 amplification cycles (94°C for 30s, 64°C for 30s, and 68°C for 1 min).
3. Conduct post-amplification melt analysis from 65°C to 95°C in 0.5°C increments every 5 seconds.
4. Determine C_q values at a threshold of 500 RFU, calculate recovered copies using a standard curve (0, 10, 10^2, 10^3, 10^4, and 10^5 copies positive control) run for each experiment, and verify that melting temperature (T_m) matches that of expected amplicon.

Table 3.1. Probe, primer, and target sequences.

Oligo	Sequence
Positive control (50 bp)	5’-CGAACCCTGCCCAGGTCGACA**CCATTCAACA**CATAGGTGAGGTCTGCTAC-3’
Reverse complement of positive control (50 bp)	5’-GTAGCAGACCTCACCTATG**TGTTGAATGG**TGTCGACCTGGGCAGGGTTCG-3’
Biotinylated probe #1 (BP1, targets positive control)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/-3’
Biotinylated probe #2 (BP2, targets reverse complement)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAACCCTGCCCAGGTCGA/3SpC3/-3’
Forward primer	5’-CGAACCCTGCCCAGGTCGA-3’
Reverse primer[a]	5’- GTA+GCAGA+CCTCACCTATGTGT-3’
Length dependence experiment	
150 bp target	5’-CGAACCCTGCCCAGGTCGACACCATTCAACACATAGGTGAGGTCTGCTACACACCATTCAATTTCATCACTGCCAATACTCCACTCTCATCTACACAACCCATTAGTACCTTACCTCGCTTCCTATCCCAATTCACTTAATCTTAAACCG-3’
80 bp target	5’-CGAACCCTGCCCAGGTCGACACCATTCAACACATAGGTGAGGTCTGCTACACACCATTCAATTTCATCACTGCCAATACT-3’
25 bp target	5’-CCGGCTGTGGGTAGCAGACCTCACC-3’
Biotinylated probe for 25 bp target	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAAGGTGAGGTCTGCTAC/3SpC3/-3’
Primers for ultrashort qPCR of 25 bp target	See Appendix A, Section 9.3.2 or Oreskovic *et al.* (2019) (127)
Comparison of single and dual biotinylated probes	
Single biotin analog of BP1	5’-/5Biosg/AAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/-3’

/52-Bio/ indicates dual biotin modification; /5Biosg/ indicates single biotin modification; /3SpC3/ indicates carbon spacer; “+G” and “+=C” indicate locked nucleic acid (LNA) bases. Target-specific probe binding sequences are underlined. A synthetic spacer region introduced to differentiate the positive control from the endogenous Mycobacterium tuberculosis complex-specific target sequence (IS6110) is bolded (see Section 3.7.4). All DNA sequences were ordered from Integrated DNA Technologies (IDT; Coralville, IA, USA) with HPLC purification.

[a] For some early experiments, an older version of the reverse primer without LNA bases was used with a PCR annealing temperature of 58°C. We later added two LNA modifications and increased the annealing temperature to 64°C to improve specificity (see Section 3.7.4).

3.5. METHODS

Experiments were carried out in urine pooled from healthy volunteers (approved by the University of Washington Human Subjects Division, IRB #48840). For all experiments, the total capture probe

concentration was kept constant at 50 pmol probe per 50 µL beads. For experiments targeting dsDNA (positive control and reverse complement), both probes BP1 and BP2 were used (25 pmol BP1 and 25 pmol BP2 per 50 µL beads) unless otherwise specified. For experiments targeting single-stranded DNA (ssDNA) (positive control only), only probe BP1 was used (50 pmol per 50 µL beads) unless otherwise specified. Some experiments used an older version of the reverse primer without LNA modifications and a lower qPCR annealing temperature of 58°C. Using these conditions, we occasionally observed late, probe-independent nonspecific amplification of residual human genomic DNA (see Figure 3.7). We updated the reverse primer design to include LNA substitutions to increase its melting temperature (T_m) to match that of the forward primer and increased the qPCR annealing temperature to 64°C (see Section 3.7.4). We found that this change minimized nonspecific amplification while maintaining high qPCR efficiency. Unless otherwise noted, we used the LNA-modified reverse primer and an annealing temperature of 64°C. For all experiments, we maintained laboratory practices intended to limit contamination, including separating pre- and post-PCR rooms, regularly decontaminating work surfaces and pipettes, using sterile filtered pipette tips, and aliquoting reagents into single-use volumes.

Percent recovery, representative calibration curves, and limit of detection. To determine the purification efficiency, 10^3 copies of 50 bp dsDNA (Table 3.1) were spiked into 10 mL urine, extracted by hybridization, and amplified by qPCR (n=6 independent experiments processed on separate days). Purification efficiency was calculated as the percent of DNA spiked into urine that was recovered and detected by qPCR. Representative calibration curves for DNA recovered by hybridization were generated by spiking 0, 5, 10, 10^2, 10^3, 10^4, or 10^5 copies (0, 0.5, 1, 10, 10^2, 10^3, or 10^4 copies/mL) of 50 bp dsDNA into 10 mL urine, extracting by hybridization, and amplifying by qPCR (n=3). Representative calibration curves for the qPCR standards were generated by adding 0, 5, 10, 10^2, 10^3, 10^4, or 10^5 copies of 50 bp dsDNA directly into qPCR. The limit of detection was verified by spiking 0 or 5 copies (0.5 copies/mL) of 50 bp dsDNA into 10 mL urine, extracting by hybridization, and amplifying by qPCR (n=6 technical replicates from the same experiment).

Length dependence. To measure the recovery across fragments of different lengths, 10^3 copies of 25, 50, 80, or 150 bp ssDNA (Table 2) were spiked into 10 mL urine and extracted by hybridization using BP1 (50, 80, 150 bp targets) or the 25 bp capture probe (25 bp target) listed in Table 3.1 (n=3 independent experiments with 3 technical replicates per experiment). The 80 bp and 150 bp fragments were amplified using the same primers and qPCR conditions as the 50 bp fragment. For this experiment, the

reverse primer without LNA modifications was used. The 25 bp fragment was amplified using ultrashort qPCR as described previously (127) and in Appendix A, Section 9.3.2.

Tolerance to varied urine conditions. To test the effects of variations expected in clinical urine samples, 10^3 copies of 50 bp ssDNA were spiked into 10 mL PBS with varied pH (5, 6, 7, 8), non-target DNA (0, 1, 10 µg sheared salmon sperm DNA), and salt (13.7, 137, 500 mM NaCl) conditions, extracted by hybridization, and amplified by qPCR using the reverse primer without LNA modifications (n=1).

Resistance to qPCR inhibition. To test for qPCR inhibition, eluate extracted from 10 mL urine (no added target) was spiked into qPCR containing 10^3 copies of 50 bp ssDNA (0, 1, 5, 10, or 20 µL eluate in 50 µL qPCR, for final 0%, 2%, 10%, 20%, or 40% eluate, respectively) (n=3 technical replicates from the same experiment). Amplification was carried out using the reverse primer without LNA modifications. An increase in quantification cycle (C_q) was used to indicate qPCR inhibition.

Comparison of single and dual biotinylated capture probes. To determine the dependence on probe concentration during immobilization, beads were functionalized with 5, 25, 50, 125, 250, or 500 pmol dual biotinylated BP1 or the analogous single biotinylated probe (Table 2) per 50 µL beads. 10^3 copies of 50 bp ssDNA were spiked into 10 mL urine, extracted by hybridization, and amplified by qPCR (n=3 independent experiments processed on separate days). To test thermostability, beads were functionalized with 50 pmol dual biotinylated BP1 or the analogous single biotinylated probe per 50 µL beads and used to extract 10^3 copies of 50 bp ssDNA spiked into 10 mL urine. Beads were either added to the sample prior to denaturation at >90°C for 15 minutes, or after the sample was denatured and cooled to room temperature (n=3 independent experiments processed on separate days). To test tolerance to free biotin, beads were functionalized with 50 pmol dual biotinylated BP1 or the analogous single biotinylated probe per 50 µL beads. Before hybridization, functionalized beads were incubated for 15 minutes at room temperature with or without 1 µM free D-biotin and washed three times with high salt wash buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20). Beads were then used to extract 10^3 copies of 50 bp ssDNA spiked into 10 mL urine (n=3 independent experiments processed on separate days). For these experiments, the reverse primer without LNA modifications was used.

Comparison of one and two probe systems. To compare recovery of dsDNA using the one and two probe systems, 1000 copies of 50 bp dsDNA were spiked into 10 mL urine, extracted by hybridization, and amplified by qPCR. For the single probe system, 50 pmol of capture probe (BP1 or BP2) was used per 50 µL beads (n=3 technical replicates from the same experiment). For the two-probe system, 25 pmol of each probe (BP1 and BP2) was used per 50 µL beads (n=3 technical replicates from the same

experiment). To confirm that introduction of a second probe had no effect on the recovery of the opposite probe, 1000 copies of 50 bp ssDNA (positive control or reverse complement) were spiked into 10 mL urine, extracted by hybridization (50 pmol BP1 or BP2 or 25 pmol each of BP1 and BP2), and amplified by qPCR (n=3 technical replicates from the same experiment). To determine if re-hybridization of complementary target strands affected recovery, 1000 copies of 50 bp ssDNA or 50 bp dsDNA were spiked into 10 mL urine, extracted by hybridization (50 pmol BP1 or 50 pmol BP2), and amplified by qPCR (n=3 technical replicates from the same experiment). For these experiments, the reverse primer without LNA modifications was used.

Statistical methods. When appropriate (i.e., when replicates were carried out across multiple, independent experiments), conditions were compared using one-way ANOVA with Tukey's honestly significant difference post hoc test. Statistical analysis was conducted using GraphPad Prism v8.1.2 (San Diego, CA, USA) with a significance level of 0.05.

3.6. Analytical characterization of sequence-specific purification method

Our sequence-specific capture method recovers 94.1% (95% CI: 82.5–105.8%) of DNA spiked into 10 mL urine (Figure 3.3A). The calculated percent recovery is occasionally >100% due to variability in qPCR quantification when using a unique standard curve for each experiment (Figure 3.4), but centers near 100% across multiple independent experiments. Figure 3.3B shows that representative calibration curves for DNA spiked into 10 mL urine and recovered by hybridization and DNA spiked directly into qPCR are visually indistinguishable, illustrating near complete recovery of DNA from urine across a range of concentrations. Using our sequence-specific capture method, we can reliably distinguish 5 copies of dsDNA spiked into 10 mL urine from negative controls (Figure 3.3C). Our protocol improves the limit of detection (≤0.5 copies/mL dsDNA, or 1 copy/mL ssDNA, approaching the detection limit of qPCR) compared to previous sequence-specific purification protocols, which had, at best, analytical sensitivities in the range of 5–100 copies/mL (Table 2.2). Our protocol also increases the input sample volume to 10 mL from a previous maximum of 1 mL (Table 2.2). Testing in our lab indicates that the increase in sample volume was crucial for detection of TB-specific cfDNA in urine from patients with pulmonary TB (Chapter 4.8).

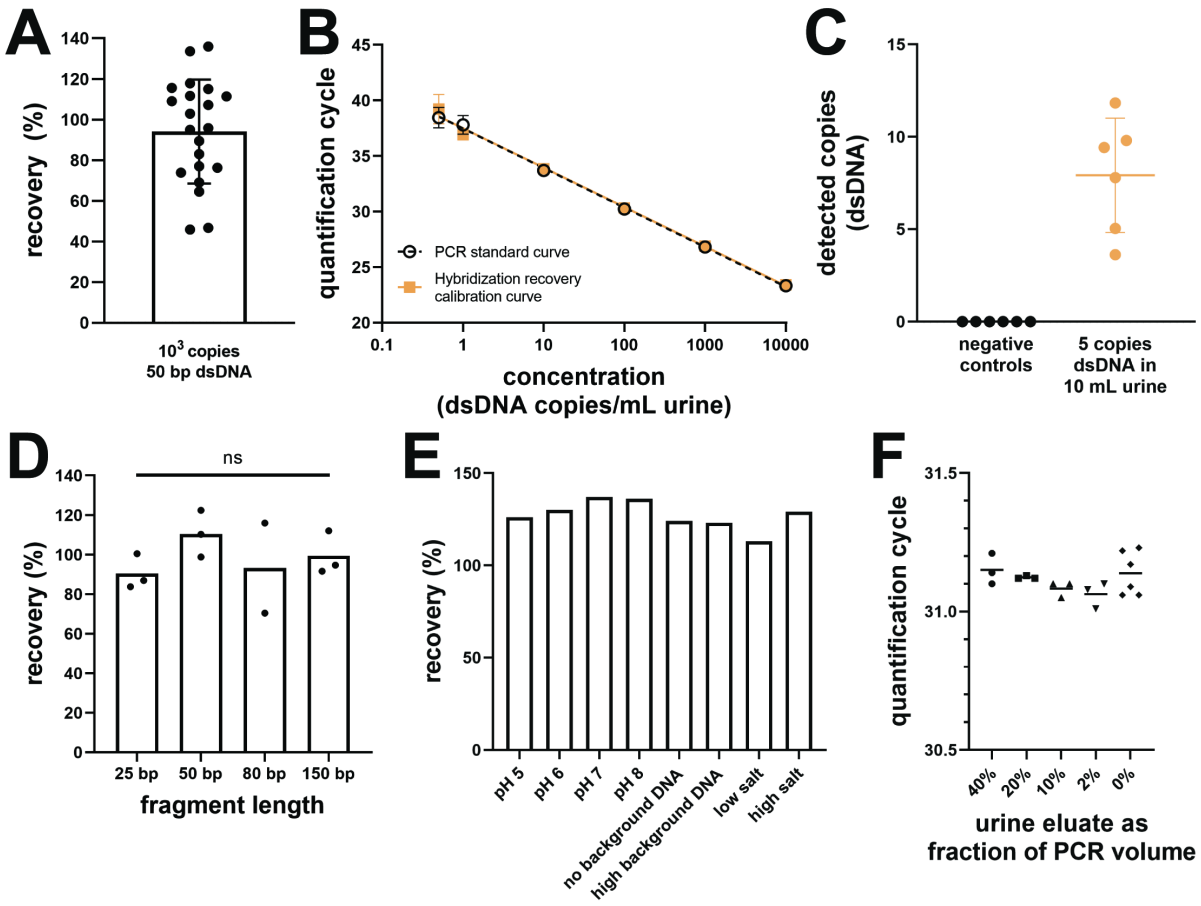


Figure 3.3: Analytical performance of the sequence-specific capture method. (A) *The sequence-specific capture method recovers nearly all TB-specific dsDNA (10^3 copies of 50 bp positive control) spiked into 10 mL urine (mean ± SD, n=21 independent experiments processed on different days).* ***(B)*** *Representative calibration curves for DNA recovered by hybridization and qPCR standards overlap across concentrations from 0.5–10^4 copies/mL 50 bp dsDNA in urine (mean ± SD, n=3).* ***(C)*** *The sequence-specific capture method reliably distinguishes 5 copies of 50 bp dsDNA spiked into 10 mL urine (0.5 copies/mL) from negative controls (mean ± SD, n=6).* ***(D)*** *Sequence-specific capture has high recovery across fragment lengths from 25–150 bp (mean, n=3 independent experiments processed on different days, 10^3 copies ssDNA input). Each dot represents the mean of 3 technical replicates performed for each independent experiment. One experiment was excluded for the 80 bp fragment length due to a calculated recovery of >200% caused by delay in amplification of the PCR standards. ns indicates not significant (one-way ANOVA with post-hoc Tukey test).* ***(E)*** *Hybridization is tolerant to variations expected in clinical urine samples, including pH (5–8), non-target DNA (0–10 µg), and salt (0–500 mM) (n=1, 10^3 copies 50 bp ssDNA input).* ***(F)*** *Sequence-specific capture enables amplification of entire output from 10 mL urine (20 µL) in a single qPCR well (50 µL) without inhibition. Eluate extracted from pooled urine (no added target) was spiked into qPCR containing a constant target concentration (10^3 copies 50 bp ssDNA). Resistance to qPCR inhibition is indicated by the lack of increase in quantification cycle (C_q) despite increasing fraction of eluate up to 40% (mean, n=3).*

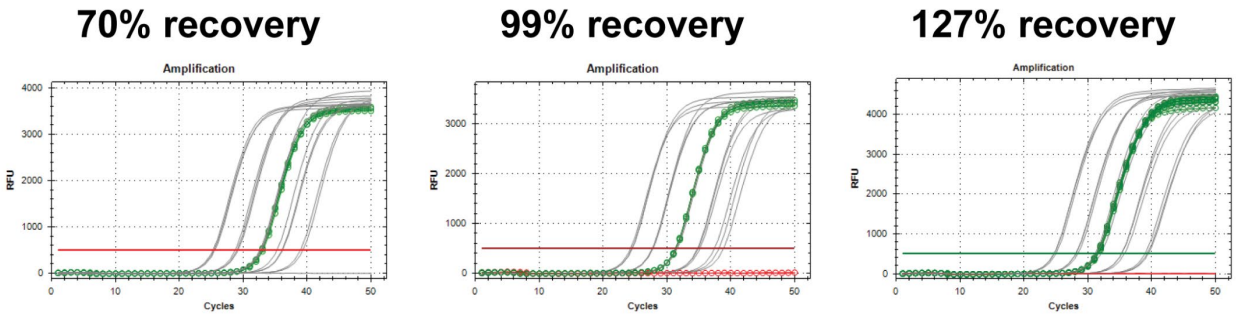


Figure 3.4: Justification for occasional >100% recovery. *In some cases, the calculated percent recovery of our assay may be >100% due to expected variations when quantifying by a qPCR standard curve. The relationship between qPCR threshold cycle and calculated starting quantity is exponential, so a small change in threshold cycle (for either the experimental sample or the qPCR standards used to generate the standard curve) can result in a large change in calculated starting quantity, and therefore percent recovery. Given here are the qPCR amplification curves for three example experiments with 70%, 99%, and 127% calculated recovery. The experimental samples (1000 copies extracted by hybridization) are shown as green curves with hollow circles. The qPCR standards (10, 10^2, 10^3, 10^4, 10^5 copies) are shown in grey and the NTCs (0 copies) are shown in red. The line at 500 RFU indicates the baseline threshold at which the threshold cycles were determined. Despite the variation in calculated percent recovery caused by minor differences in the standard curves, the 1000 copy hybridization output visually overlaps with the 1000 copy PCR standard across all three experiments, indicating near complete recovery of DNA by sequence-specific capture. We run a new qPCR standard curve for every experiment, so while calculated percent recovery may diverge from 100% for technical replicates from an individual experiment, we expect the calculated percent recovery across multiple independent experiments to center around 100%, as seen in Figure 3.3A. Minor differences in calculated percent recovery for experiments with only technical replicates from the same experiment and/or quantified based on different standard curves are not necessarily meaningful and should be interpreted with this in mind.*

Critically for urine cfDNA, which is made up primarily of short fragments (majority <100 bp) (45, 54–56), our sequence-specific capture method is not dependent on DNA fragment length and maintains high recovery across fragments from 25–150 bp (Figure 3.3D). Combined with an ultrashort stem-loop qPCR design that I developed (127), we can capture and amplify fragments as short as 25 bp with high efficiency (84% recovery). In contrast, conventional DNA extraction based on adsorption of DNA to silica under chaotropic conditions is dependent on DNA fragment length, with reduced recovery of shorter fragments. DNA adsorption to silica is driven by hydrophobic interactions between dehydrated silica and DNA surfaces and hydrogen bonding between silica and the DNA backbone (74), both of which are dependent on DNA fragment length. We previously reported that several silica-based methods for urine cfDNA extraction, both published lab-based protocols and commercially-available kits, have reduced recovery of short DNA fragments (127). We determined that the frequently-used Wizard silica resin method, which was used in the first study demonstrating the presence of cfDNA in urine (3), had low recovery (<35%) of 40–150 bp fragments, with a further drop in recovery for 25 bp fragments (<2%)

(127). Even the best-performing method that we identified (Q Sepharose anion exchange resin (79)), which had moderate recovery (63–75%) of 40–150 bp fragments, had low recovery of the shortest 25 bp fragment (9%) (127). For the full comparison of the analytical performance of urine cfDNA extraction methods, see Chapter 4.

Sequence-specific capture is expected to perform well across varied clinical urine samples due to the insensitivity of DNA hybridization to properties that vary in urine samples. Specifically, we have demonstrated tolerance to a wide range of pH (5–8), non-target DNA (0–10 μg), and salt (0–500 mM) conditions (Figure 3.3E). We expect clinical urine samples to be highly variable, with pH from 5–8 (mean 5.99–6.43) (131), background DNA from <1–200 ng/mL (3, 47, 48), and salinity from 11–416 mM (mean 97–148 mM) (132). Particularly notable is sequence-specific capture's robust performance for low concentrations of non-target background DNA, as total urine cfDNA concentrations are expected to be low (ng/mL) (3, 47, 48). In contrast, silica-based DNA extraction may perform poorly for low DNA concentrations, which is why many silica-based kits require the addition of carrier nucleic acids to maintain high yields for low-concentration samples. In the presence of high chaotrope concentrations, a fraction of DNA may bind irreversibly to the silica surface due to strong hydrophobic interactions (75, 76). For high-concentration samples, this loss may be trivial, but for low-concentration samples like urine cfDNA the unrecoverable DNA may make up a significant fraction of the input (75, 76). In our previous comparison of urine cfDNA extraction methods (Chapter 4), we found that the Wizard silica resin method was highly dependent on sample composition, and was thus likely to fail in a subset of clinical urine samples with high pH and/or low non-target DNA concentrations (127).

Our sequence-specific capture protocol also efficiently removes qPCR inhibitors in urine (urea, creatinine, etc.) that may affect downstream amplification, enabling amplification of cfDNA from 10 mL healthy control urine in a single qPCR well without inhibition (500X concentration, Figure 3.3F). Avoiding dilution of eluate during PCR will help maximize sensitivity for detection of low concentration samples. Unfortunately, I found that PCR inhibition did occasionally occur when testing urine samples from hospital inpatients, likely in part due to transfer of small amounts of magnetic beads to PCR, often in samples with blood and/or proteinuria (see Section 7.6 for more details and suggested addition of a Proteinase K digestion step to mitigate effects of sample variability).

Although not a necessity for dilute urine cfDNA, sequence-specific capture is also beneficial for samples where high concentrations of co-extracted non-target DNA can act as an inhibitor of downstream amplification (e.g., stool, sputum, whole blood) (109, 113, 124). Non-target DNA amounts of 2–10 μg or

higher have been reported to inhibit amplification (109, 113, 133), but can be effectively removed by hybridization (Figure 3.3E). While we have not tested sample types besides urine here, we anticipate that our approach could also be broadly useful for samples with a high background of non-target DNA.

3.7. KEY DESIGN FEATURES FOR MAXIMUM SENSITIVITY SEQUENCE-SPECIFIC CAPTURE

During the assay development process, I iterated through several variations of the sequence-specific capture method. In this section I describe the most important design features that ultimately led to the strong analytical and clinical performance of the resulting TB urine cfDNA assay. More details regarding optimization efforts and other variables that had a minor (or no) effect are given in Appendix B.

3.7.1. *Increased input sample volume using surface-based hybridization approach*

3.7.1.1. Solution-based hybridization capture assay (1 mL)

I initially used a solution-based hybridization approach (capture probes hybridized to target prior to immobilization on magnetic beads) due to the potential for faster hybridization kinetics in solution compared to on a bead surface (134). When I first began work on this project, the capture protocol had low sensitivity, limited sample volume, high hands-on time, and high cost (Table 3.2). My first objectives were to decrease the cost and assay time sufficiently to enable higher throughput of assay development experiments and to increase the sample volume to at least 1 mL. I achieved these goals primarily by reducing the required amount of magnetic beads, which are by far the most expensive component of the assay. I also reduced the target footprint from 129 bp to 40 bp and reduced the elution volume from 80 µL to 20 µL. The improved solution-based capture method had high recovery (73–84% of 25–150 bp DNA) and was reliably capable of detecting down to 10 copies of spiked TB DNA in 1 mL urine (see Chapter 4 for full analytical characterization). The protocol is given in Appendix A, Section 9.2.1. Despite rigorous optimization efforts and the strong analytical performance of the solution-based capture assay, I was not able to detect any TB cfDNA in a subset of 1 mL clinical urine samples from the TB Control Program (Public Health – Seattle & King Co.) using this protocol. Because capture probes were hybridized to target cfDNA prior to immobilization on beads, an excess of beads was required for the solution-based method to avoid competition of probes with endogenous biotin in urine. The beads make up much of the cost of the assay (~$15/sample for beads *vs.* ~$0.50/sample for remaining reagents), so simply scaling up solution-based hybridization from 1 mL to a larger 10 mL urine sample was cost prohibitive. I attempted to further reduce the amount of beads used or switch to alternate, cheaper beads, but both caused recovery to drop (see Appendix B, Figure 10.2 and Figure 10.3).

Table 3.2: Summary of improvements made to sequence-specific capture method

	Solution-based hybridization (original protocol)	Solution-based hybridization (optimized protocol)	Surface-based hybridization (final protocol)
Urine volume	75 µL	1 mL	10 mL
Cost per sample	$100	$15	$9
Cost per mL	$100	$15	$0.90
Minimum reliably detected concentration	100 copies/mL	≤10 copies/mL	≤0.5 copies/mL
% Recovery (approximate)	<10%	70–90%	>90%
Minimum target footprint	129 bp	25–50 bp	25–50 bp
Total purification time	8+ hr	2–3 hr	2–3 hr
Output	80 µL	20 µL (into single PCR well)	20 µL (into single PCR well)
Concentration factor	0.9375X	50X	500X
Success in clinical samples?	No	No	Yes

3.7.1.2. Surface-based hybridization capture (10 mL)

Switching to a surface-based hybridization approach enabled me to avoid competition with endogenous biotin by pre-immobilizing probes on beads prior to hybridization, reduce bead consumption and cost per milliliter of sample analyzed ($0.95/mL), and scale up the reaction volume to 10 mL. Without increasing the amount of beads used (keeping the cost per sample the same), the sample volume can be further scaled up to 20 mL with minimal to no loss in performance (80–100% recovery) and to 30–40 mL with only moderate loss in performance (50–60% recovery) (see Section 7.5). While the kinetics of surface-based hybridization are slower than the kinetics of solution-based hybridization, using microparticles rather than planar surfaces helps to mitigate the effect on reaction rate. The rate constant for surface-based hybridization on microparticles with optimal probe density has been reported as only 5-fold lower than for solution-based hybridization (135). With my optimized protocol, surface-based hybridization remains rapid, with near complete hybridization in ≤30 minutes at room temperature. Increasing the hybridization temperature or time does not lead to an improvement in performance (Appendix B, Figure 10.4).

Increasing the sample volume from 1 mL to 10 mL was critical for successful detection of TB cfDNA in clinical urine samples. The percent recovery of spiked DNA from 10 mL urine using surface-based hybridization was not initially as high as that from 1 mL urine using solution-based hybridization, but I

identified several key parameters, particularly capture probe density on the bead surface (Section 3.7.2), that were critical for efficient hybridization when probes are pre-immobilized on beads. After optimization of the surface-based hybridization method, its recovery is as high as that of solution-based hybridization, and it improves the limit of detection by 10-fold compared to solution-based hybridization (Table 3.2). The protocol is given in Section 3.4.

3.7.1.3. Comparing surface-based and solution-based hybridization capture

Based on my experience, I generally recommend surface-based hybridization over solution-based hybridization, particularly when large sample volumes are required. Surface-based hybridization with dual biotinylated capture probes (Section 3.7.2) eliminates the risk of biotin interference while remaining cost-effective for large sample volumes, while solution-based hybridization may be compromised by high endogenous biotin concentrations and is more expensive. Assay time and percent recovery using the two approaches are comparable, as long as conditions are optimized. Specifically, optimal probe density is critical for success with surface-based hybridization, as discussed below.

3.7.1.4. Surface-based hybridization capture with covalent coupling (10 mL)

To reduce cost (approximately $9/sample for surface-based hybridization using a biotin/streptavidin system), I investigated using a covalently coupled bead/probe system instead (approximately $4/sample). It is also possible that using beads without a protein coating may improve bead behavior in variable clinical samples (particularly those with blood and/or proteinuria, see Section 7.6). I chose Dynabeads MyOne Carboxylic Acid due to their similarity (size and hydrophilicity) to the current streptavidin beads. I designed an amine-modified probe (/5AmMC6/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/) that can be covalently coupled to the bead surface through a one- or two-step carbodiimide crosslinking reaction. I had more success with one-step EDC coupling than with two step EDC-NHS coupling. The protocol I used for bead functionalization using one-step EDC coupling is given in Appendix A, Section 9.2.4.

Unfortunately, recovery using covalently functionalized beads was at most 30–50% in my preliminary testing. Due to the instability of the activated ester intermediate during carbodiimide crosslinking, a high concentration of amine probe must be used to ensure high coupling efficiency, making it difficult to finely control probe density and distribution on the bead surface. If the probe concentration is too low, the coupling efficiency will be low and not enough probes will be immobilized for efficient hybridization. If the probe concentration is high, coupling efficiency will be high but the probe density will be too high, leading to steric hindrance and low hybridization recovery. I deprioritized this work and instead focused

on clinical testing using the better-performing biotin/streptavidin system. If a covalently coupled system is desired, future work could include addition of a sacrificial amine molecule such as amine-PEG during coupling to decrease the probe density while keeping total amine concentration and coupling efficiency high. PEGylation would also help to passivate the bead surface and reduce nonspecific, potentially irreversible binding of DNA. Alternate coupling chemistries with more stable intermediates, such as tosylactivated beads, could also enable use of lower probe concentrations during coupling. Finally, a method to directly measure probe coupling efficiency rather than using hybridization recovery as a surrogate measure would greatly ease optimization of coupling conditions.

3.7.2. *Dual biotinylated capture probes moderate probe density, improve thermostability, and enhance tolerance to endogenous biotin*

Dual biotinylated probes, which have two biotin molecules in series at the 5' end of the probes, are a key design element of the sequence-specific purification protocol that help ensure its high sensitivity and robust performance. Switching from single to dual biotinylated probes was one of the most critical turning points that led to an increase from 50–60% recovery (early version of protocol) to near 100% recovery when using a surface-based hybridization format. As shown in Figure 3.5A, dual biotinylated probes increase recovery and reduce reliance on the probe concentration used during probe immobilization compared to single biotinylated probes. When I varied the amount of probes added to beads during the probe immobilization step, single biotinylated probes had only a narrow peak in probe concentration (25–50 pmol probe per 50 µL beads) that led to maximum recovery. Higher probe concentrations resulted in reduced recovery, with 50% or less recovery for probe concentrations at or above that expected to saturate beads (estimated as 250 pmol per 50 µL beads for single biotinylated probes). Dual biotinylated probes, on the other hand, had high recovery across a broad range of probe concentrations (25–500 pmol probe per 50 µL beads). Even when dual biotinylated probes were added at concentrations beyond that expected to saturate beads (estimated as 125 pmol per 50 µL beads for dual biotinylated probes), recovery remained near 100%.

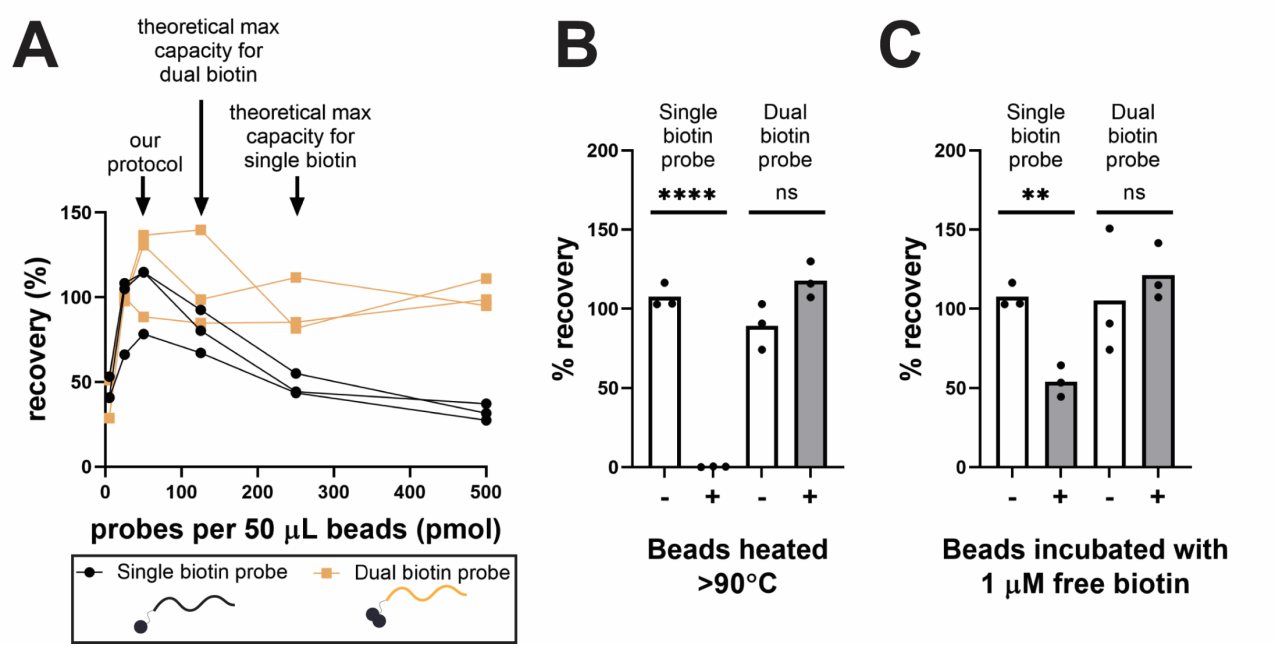


Figure 3.5: Dual biotinylated capture probes improve hybridization compared to single biotinylated capture probes. (A) *Dual biotinylated probes (in orange) facilitate optimal spacing and density of probes on the bead surface, increasing recovery and reducing reliance on the probe concentration used during bead functionalization compared to single biotinylated probes (in black) (n=3 independent experiments on different days; 10^3 copies 50 bp ssDNA input).* ***(B)*** *Beads functionalized with dual biotinylated probes are thermostable up to at least 90°C, while beads functionalized with single biotinylated probes lose function after heating to 90°C. Grey columns labeled as "+" were heated to >90°C for 15 minutes prior to hybridization; white columns labeled as "-" were unheated controls (mean, n=3 independent experiments on different days; 10^3 copies 50 bp ssDNA input).* ***(C)*** *Beads functionalized with dual biotinylated probes tolerate incubation with free biotin, while beads functionalized with single biotinylated probes have reduced recovery after incubation with free biotin. Grey columns labeled as "+" were incubated for 15 minutes with 1 µM free biotin; white columns labeled as "-" were controls not incubated with free biotin (mean, n=3 independent experiments on different days; 10^3 copies 50 bp ssDNA input). ** indicates P value of 0.001 to 0.01, **** indicates P value of <0.0001, ns indicates not significant (one-way ANOVA with post-hoc Tukey test).*

Probe density has a significant, well-documented impact on the extent of DNA hybridization on surfaces, due to both thermodynamic and kinetic effects (136, 137). If the probe density is too low, the kinetics of hybridization will be slow, resulting in low recovery, as is likely the case in Figure 3.5A for both single and dual biotinylated probes at 5 pmol per 50 µL beads. If the probe density is too high, steric hindrance and electrostatic repulsion near the bead surface can reduce both the rate of hybridization and the final hybridization efficiency (136, 137), as is likely the case in Figure 3.5A for single biotinylated probes at ≥125 pmol probe per 50 µL beads. Several studies have observed a similar density-dependent peak in hybridization efficiency for immobilized probes (136, 138–140), including on magnetic beads (135), with hybridization efficiency reported to drop at probe densities greater than 4–5 x 10^{12} probes/cm^2 (136,

138). I selected 50 pmol dual biotinylated probe per 50 µL beads as the concentration to use during coupling, which corresponds to a probe density of 1.9–2.7 x 10^{12} probes/cm^2 (370 – 526 $Å^2$/probe) on the bead surface, within the optimal range for rapid, efficient hybridization. Importantly, heterogeneity of probe distribution across the surface can also lead to sub-optimal recovery (141). Dual biotin modification of capture probes likely helps avoid steric hindrance both by limiting overall maximum probe density (by saturating streptavidin at a lower probe density) and facilitating optimal spatial distribution of probes on the bead surface (by limiting to only one or two probes per tetrameric streptavidin molecule). The practical implication of this phenomenon is that non-optimal bead:probe ratios during probe immobilization are unlikely to compromise assay performance if dual biotinylated probes are used, eliminating the need for careful tuning of probe density and reducing variability between bead batches.

I found only a single instance of dual biotinylated probes used previously for hybridization-based sample preparation, which, although overall recoveries were lower (<30%), reported a similar effect for reduced dependence on probe concentration relative to single biotinylated probes (120). The authors measured dual biotinylated probe density at saturation and confirmed that it was half of that of single biotinylated probe density at saturation, as expected (120).

In addition to moderating probe density, the dual biotin modification is advantageous because it increases thermostability of the bead-probe linkage. Beads functionalized with dual biotinylated probes maintained recovery after heating at >90°C for 15 minutes, while beads functionalized with single biotinylated probes lost all function after heating (Figure 3.5B). Dual biotinylated probes likely increase biotin-streptavidin thermostability by more fully saturating streptavidin with biotin while still maintaining an optimal probe density. Saturation with biotin increases the T_m of thermally-induced denaturation of streptavidin from 75°C to 112°C (142). The increase in thermostability is incremental, with streptavidin 50% saturated with biotin showing two distinct thermogram peaks (142). Dual biotinylated probes have been used previously to improve biotin-streptavidin thermostability during heat elution (120, 143) and enable PCR thermocycling without leeching of biotinylated DNA from streptavidin beads (120, 144). For our purposes, enhanced streptavidin thermostability means that beads can be added to urine before the denaturation step without loss in performance. Denaturation of beads functionalized with single biotinylated probes, or adding beads functionalized with single biotinylated probes to heated samples immediately after denaturation prior to cooling, results in reduced recovery due to loss in function as streptavidin denatures. Cold shock, which is commonly used

to cool denatured samples while minimizing DNA renaturation, could avoid this problem for single biotinylated probes but is not effective for large volume samples. Allowing for pre-heating of beads may also improve availability of surface-immobilized probes, leading to an increase in hybridization recovery (136).

Saturating unbound streptavidin with biotin after probe immobilization may achieve a similar improvement in thermostability, but single biotinylated probes are negatively affected by the presence of high biotin concentrations. Although the streptavidin-biotin association is strong, some dissociation does occur (145). In the presence of high concentrations of free biotin, competitive dissociation may displace bound biotinylated probes over time. I observed reduced recovery after incubating beads functionalized with single biotinylated probes with 1 µM free biotin for 15 minutes, but when using dual biotinylated probes, the difference in recovery with and without biotin incubation disappeared (Figure 3.5C). Endogenous biotin is a known interferent in streptavidin-based diagnostic assays (146), so the tolerance of dual biotinylated probes to high concentrations of free biotin also eliminates the risk of compromised results in urine samples with high endogenous biotin concentrations. Covalently immobilized probes may also improve thermostability and tolerance to endogenous biotin, but it is difficult to tailor the probe density and spacing during the conjugation reaction. Using dual biotinylated probes with streptavidin-coated magnetic beads retains the ability to carefully control probe density, one of the key advantages of biotin-streptavidin immobilization over covalent conjugation, while also maintaining thermostability and tolerance to endogenous biotin, characteristics which are lacking for single biotinylated probes.

3.7.3. *Two probe system enables recovery of both strands of double-stranded DNA*

Because hybridization targets single-stranded DNA, it typically only captures half of available target molecules (i.e., DNA strands) compared to non-specific extraction methods such as silica adsorption that capture both strands of double-stranded DNA. To overcome this limitation, I designed two capture probes, one complementary to each strand of the dsDNA target (Figure 3.6A-B). As shown in Figure 3.6C, the two-probe system increases detected copies by approximately two-fold compared to either single probe alone by enabling recovery of both strands of dsDNA. This design element is especially important for low-concentration targets where the number of purified target copies may be near the limit of detection of qPCR. By recovering both strands of the dsDNA target, sensitivity improves by doubling the number of potential amplicons for qPCR and increasing the likelihood of successful detection in low-concentration clinical samples. The addition of a second capture probe does not affect

recovery of the first probe when the total probe concentration is held constant (Figure 3.6D). Re-hybridization of dsDNA target stands is thermodynamically favored over hybridization to capture probes, but the high concentration of capture probes kinetically favors probe-target hybridization. Re-hybridization of dsDNA is negligible and does not affect target recovery after 30 minutes of hybridization at room temperature (Figure 3.6E).

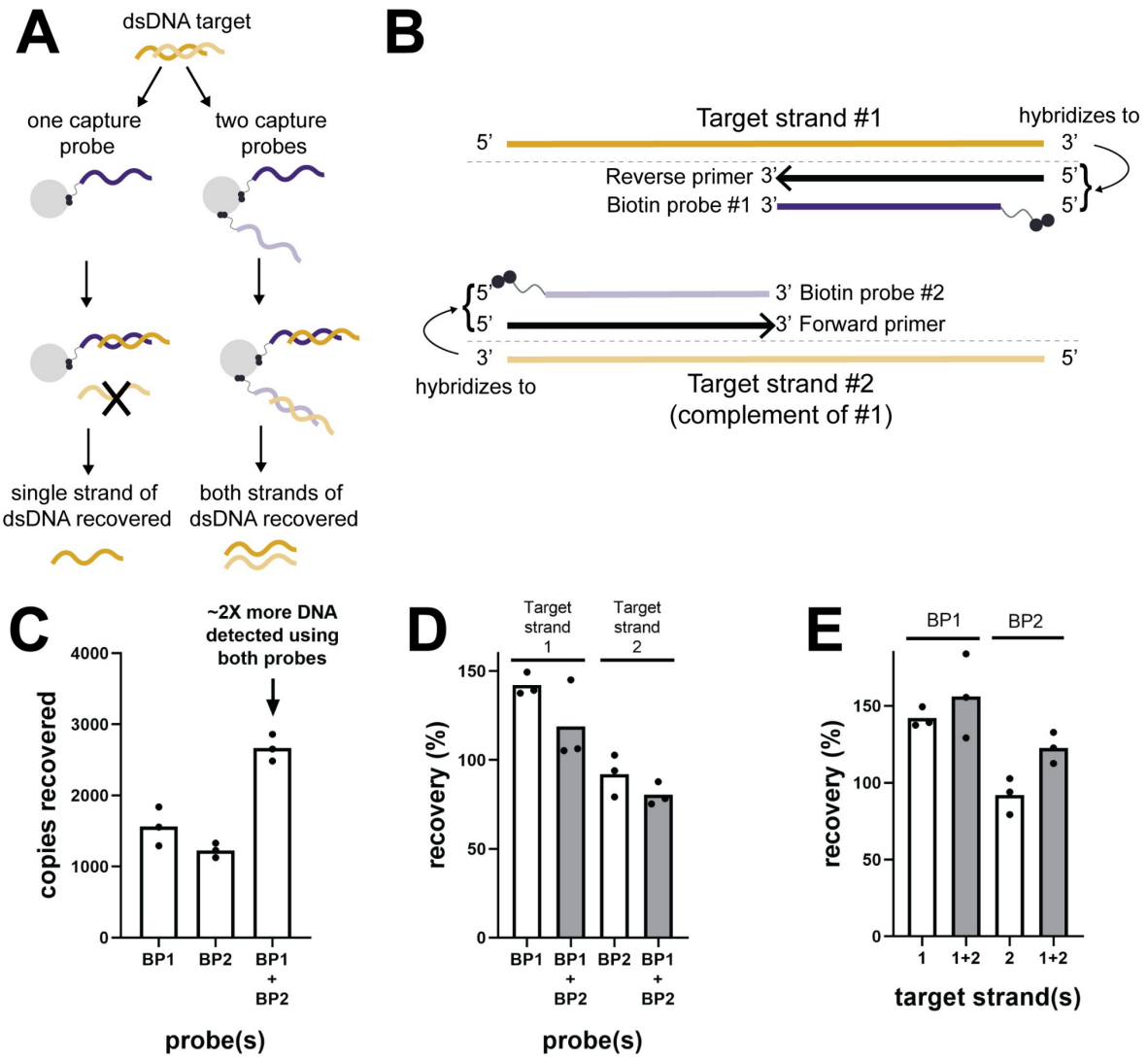


Figure 3.6: Two probe system enables recovery of both strands of double-stranded DNA. (A) *Rationale for two-probe system. When using a single capture probe (left column), only one strand of the dsDNA target is recovered by hybridization; the other strand is discarded. When a second capture probe is added to target the complementary strand (right column), both strands of dsDNA are recovered, doubling the concentration of detectable DNA.* ***(B)*** *Schematic of two-probe system design. To minimize target footprint, probes are truncated versions of the forward and reverse qPCR primers (15 – 20 bp). Probes hybridize to different sub-regions of the target sequence so that opposite probes do not hybridize to each other.* ***(C)*** *Using both capture probes results in an approximate 2-fold increase in detected dsDNA compared to either single probe (mean, n=3 technical replicates from the same experiment; 10^3 copies 50*

bp dsDNA input). Each target strand was quantified by a different standard curve, so the minor difference observed in calculated percent recovery for BP1 and BP2 is unlikely to be meaningful (see Figure 3.4*; same applies for panels D-E).* ***(D)*** *Addition of a second probe does not affect recovery of ssDNA by the opposite probe (mean, n=3 technical replicates from the same experiment; 10^3 copies 50 bp ssDNA input).* ***(E)*** *Re-hybridization of complementary dsDNA target strands is negligible compared to probe-target hybridization and does not affect recovery (mean, n=3 technical replicates from the same experiment; 10^3 copies 50 bp ssDNA dsDNA input). Because plots in this figure show technical replicates from the same experiment, the data are not independent and statistical analysis was not conducted.*

3.7.4. *LNA-substituted primers to improve specificity*

There are two primary sources of false positives in diagnostic tests: nonspecific amplification and contamination. To distinguish true positives from false positives due to nonspecific amplification (either off-target amplification or primer dimer amplification), sequence-specific fluorescent probes such as molecular beacons or TaqMan probes are often used during qPCR. These probes are designed to bind to a region between the primers that is specific to the target of interest. Amplicon length must be minimized to the combined length of the two primers when targeting short urine cfDNA, leaving no room for a sequence-specific detection probe (Figure 3.7A). Furthermore, the amplicon length is similar to that which could arise from nonspecific amplification of a primer dimer. Post-amplification analysis methods that examine amplicon length including melt curve analysis and gel electrophoresis would not be capable of distinguishing between target and primer dimer amplification.

To keep the amplicon length as short as possible while avoiding false positives due to nonspecific amplification, I relied on careful primer design and *in silico* analysis. I screened candidate primers using NCBI Primer BLAST to ensure that there were no amplifiable products from the human genome or other organisms other than the MTB complex. I tested the candidate primers and selected the pair with the most resistance to no template control (NTC) amplification, which resulted in no NTC amplification up to at least 45 cycles. In clinical samples, however, the selected primer pair occasionally amplified a nonspecific product with high T_m (78–80°C). I discovered that spiking 10 ng of human genomic DNA (Promega) into PCR replicated the occasional high-T_m product observed in clinical urine samples. Although sequence-specific purification removes most non-target DNA, minor non-specific adsorption of non-target DNA to the bead surface (found to be independent of whether or not beads were functionalized with capture probes) results in some residual non-target DNA, so well-designed primers remain essential. I was unable to completely eliminate nonspecific adsorption by blocking beads with BSA, Denhardt's solution, or sheared salmon sperm DNA, or by switching to beads with a different surface chemistry (Dynabeads MyOne Streptavidin T1). For most applications, late, infrequent nonspecific amplification would not be much of a concern, especially when it can be distinguished from

target-specific amplification by post-amplification melt curve analysis. However, when attempting to detect such low concentrations of cfDNA, with the ultimate goal of single copy sensitivity, there is a possibility that nonspecific amplification could compete with target-specific amplification and reduce sensitivity.

To improve primer specificity, I used locked nucleic acids (LNA). LNA nucleotides are a modified version of RNA nucleotides with a methylene bridge between the 2' oxygen and 4' carbon to fix the ribose monomer in the 3'-endo conformation, encouraging formation of an A-form helix rather than a B-form helix (147). LNA substitutions within DNA oligonucleotides increase duplex stability and improve specificity (147). They have previously been demonstrated to improve the mismatch discrimination of hybridization probes (148) and the yield and maximum annealing temperature of PCR primers (149, 150). To discourage nonspecific amplification, I used LNA substitutions to better match the forward and reverse primer T_ms (Figure 3.7B) and increased the PCR annealing temperature to slightly above the empirically-measured primer T_m (from 58°C to 64°C), which improved specificity (no amplification of 10 ng human genomic DNA up to 45 cycles) and PCR efficiency (from low 90s to >97%) (Figure 3.7C). When testing clinical samples (Chapter 4.8), this optimized primer set and PCR cycling conditions resulted in almost no high-T_m nonspecific amplification (4/222 [1.8%] of samples) and 100% specificity.

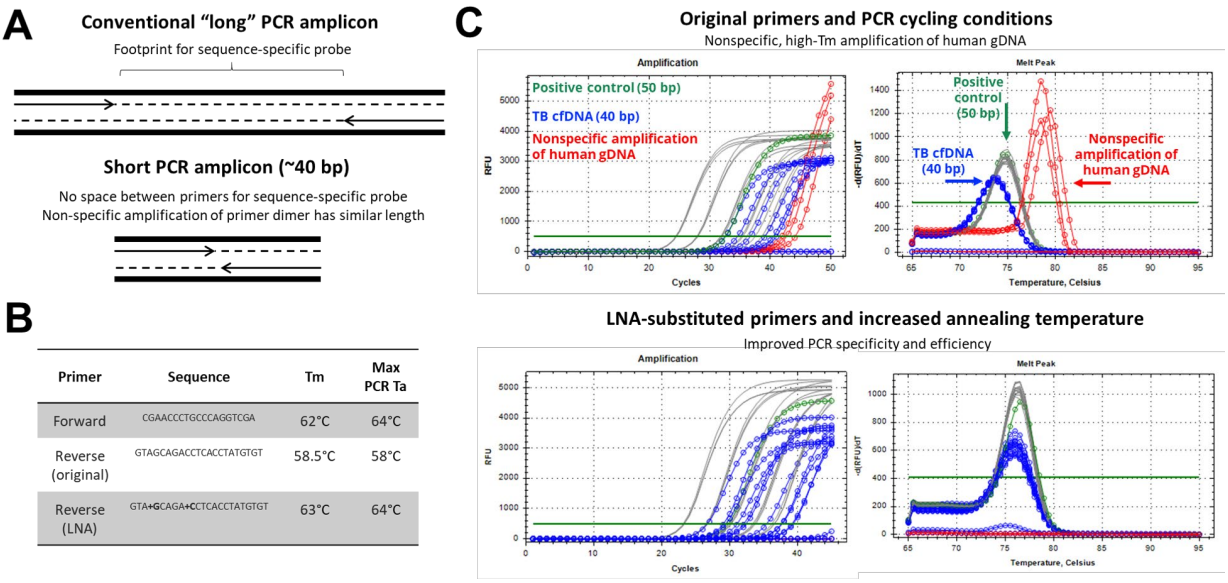


Figure 3.7: Designs to ensure specificity of short PCR. (A) *Short PCR design has no footprint between primer sites for binding of a sequence-specific probe. Amplicon is limited to the combined length of the two primers, so nonspecific amplification of a primer dimer would have a similar length and may be indistinguishable by melt analysis.* ***(B)*** *LNA substitutions can be used to carefully match primer melt temperatures.* ***(C)*** *LNA-substituted primers and an elevated PCR annealing temperature reduce*

nonspecific amplification of 10 ng human genomic DNA (Promega, red) without reducing PCR efficiency. Melt analysis can distinguish between endogenous TB amplicon (blue) and synthetic positive control (green), preventing false positives due to contamination with positive control sequence.

Using LNA-substituted primers may be a valuable tool for improving the performance (efficiency and/or specificity) of existing primer sets, particularly when primer design space is limited, target sequence is non-optimal for primer design (e.g., AT-rich regions), or other oligos such as probes would be cumbersome or expensive to redesign. For cfDNA in particular, designing optimal primers can be challenging because the amplicon length is fixed, with primers directly adjacent to one another, limiting permutations during the generation of initial primer candidates. LNA substitutions may help improve non-optimal primer candidates. LNA substitutions could also be a useful tool for better matching primer T_ms during the design of multiplexed PCR or reducing primer length to amplify shorter targets (149, 150).

Guidelines for designing LNA-substituted primers:

1. Empirically measure the T_m of original primers to their reverse complement sequences under PCR conditions.
    - Anneal from 95°C to 45°C, decreasing by 1°C every 5 seconds.
    - Melt from 45°C to 80°C, increasing by 0.5°C and reading the plate every 5 seconds.
2. Design LNA-substituted versions of the primer with lower T_m so that it matches the primer with higher T_m.
    - Estimates for T_m of LNA-substituted primers can be calculated using IDT OligoAnalyzer (LNA bases indicated as "+X"). While the calculated T_ms may not match the empirically measured T_ms, you can determine a target calculated T_m for the LNA-substituted primer by adding the difference of the two empirically-measured primer T_ms to the calculated T_m of the original (lower T_m) primer.
    - Based on my experience, I recommend substituting 1-3 non-consecutive bases within the central to 5' half of the primer. LNA substitutions at the 3' end inhibited amplification. Both of these observations are consistent with previous reports (149, 150).
3. Re-measure the T_m of the LNA-substituted primer using its reverse complement, as above.
4. Conduct an annealing temperature gradient covering temperatures just at or slightly above the new empirically measured T_m to find the maximum annealing temperature that does not result in a drop in PCR efficiency. In my experience, PCR annealing temperatures of 0–2°C above the lowest primer T_m are optimal.

The second possible cause of false positives is contamination of samples with the template sequence. This problem is not specific to the short target used here but is of particular importance due to the low concentrations of TB cfDNA we need to detect (possibly down to a single copy). By minimizing the likelihood of false positives due to contamination, we can more confidently detect true positives even with very low amounts of TB cfDNA. In addition to the typical engineering controls to prevent contamination (e.g., separate pre- and post-PCR spaces, single-use reagent aliquots, filtered pipette tips, regular decontamination of surfaces and pipettes), I designed synthetic positive controls that are distinguishable from the native TB target sequence by post-amplification melt analysis. The synthetic positive controls are amplified by the same primers but contain a 10 bp spacer region between the primer binding sites, increasing the total amplicon length to 50 bp and raising the melting temperature by approximately 1°C compared to the native 40 bp TB amplicon. In the PCR buffer used here, the endogenous TB target sequence melts at 75–76°C, while the synthetic positive control melts at 76.5°C (Figure 3.7C). This synthetic positive control can be used for PCR standard curves and as a spike-in target for assay development experiments and clinical testing to reduce the risk of false positives due to contamination. False positives not distinguishable by melt analysis could still potentially arise from cross-contamination of clinical samples (from a true positive sample to a negative sample) or from amplification of primer dimers (rare and infrequent using the optimized PCR primers and cycling).

3.8. Designing capture probes for application to new targets

Designing capture probes for new targets is straightforward. To reduce target footprint and minimize design effort, I recommend using truncated versions of the qPCR primers (see Figure 3.6B). This approach will also automatically limit secondary structure, which should be minimal in well-designed primers. Nucleotides should be removed from the 5’ end of the primers until a length of 15–20 bp is reached and the melting temperatures of the two probes are similar (as a rule of thumb, ~60°C as calculated in hybridization conditions of 1 M NaCl and 2 nM probe). In addition to the target-specific binding region, probes should contain a 5’ dual biotin modification, a 20 bp poly(A) spacer between the biotin modification and the target-specific binding region, and a 3’ carbon spacer, as described in Table 3.1 and Figure 3.2A. Capture probes should be ordered with HPLC purification to ensure full-length probes. The sequence-specific capture method should have high, robust recovery independent of probe sequence. In our experience, implementation of new capture probes does not require design iterations or assay optimization. For example, by following the above design rules, I designed capture probes for three additional multicopy elements within the TB genome that had 66–109% recovery (see Section 7.3).

In general, existing qPCR assays should work in combination with hybridization capture. For urine cfDNA, it is important to keep the qPCR amplicon as short as possible (ideally ≤50 bp) to maximize the fraction of cfDNA fragments that contain both primer sites and can be successfully amplified. Because the exact eluate volume varies slightly across samples, I found that partial neutralization leaving the eluate slightly basic had the least effect on downstream amplification. In optimization experiments where I controlled the eluate pH, the qPCR tolerated slightly basic eluate better than slightly acidic eluate. For new targets, I recommend using qPCR conditions with high buffering capacity (for example, 80 mM Tris-SO_4 used here). If amplification inhibition is observed, I suggest supplementing the qPCR buffer with additional Tris-HCl. To easily identify contamination and prevent false positives during clinical testing, I also recommend adding a 10–20 bp spacer region between the primer binding sites on your target of interest to design a synthetic positive control that can be distinguished from the native target of interest by post-amplification melt analysis (see Section 3.7.4).

To summarize, the suggested selection criteria for target, primer, and probe sequences are as follows:

- The target sequence should be selected for specificity to the target of interest. Targeting multi-copy genomic elements may help improve clinical detection sensitivity.
- The primers should be designed to amplify as short as possible (40–50 bp maximum for urine cfDNA) of a region within the target sequence while maintaining specificity.
- The capture probe sequences should be truncated versions of the primer sequences, with bases removed from the 5' end to reach lengths of 15–20 bp long and melting temperatures of ~60°C at a concentration of 2 nM probe in 1 mM NaCl. A dual biotin modification and 20 bp poly(A) spacer should be added to the 5' ends, and a C3 spacer should be added to the 3' ends.

3.9. Potential applications, limitations, and future work

I developed this sequence-specific capture protocol with the goal of improving the sensitivity of TB diagnosis from urine cfDNA, but we expect that it will be versatile across urine cfDNA applications such as infectious disease diagnosis (4, 42), cancer detection (3, 38, 126), organ transplant monitoring (43, 44), and fetal genetic screening (3, 45). It may also be broadly useful for targeting cfDNA in other sample types (e.g., plasma, cerebral spinal fluid, saliva) or for applications beyond cfDNA. Sequence-specific capture is likely to be especially beneficial for removing high concentrations of non-target background DNA, which can inhibit downstream amplification and contribute to non-specific amplification (109, 113, 124), and for concentrating and detecting dilute targets from large volume samples. Potential

applications that may benefit from the sequence-specific nature of hybridization include infectious disease diagnosis from whole blood, sputum, and stool, all of which contain high concentrations of amplification inhibitors, including host DNA, and environmental monitoring, which faces similar challenges of dilute targets in large sample volumes with an excess of non-target DNA. Sequence-specific hybridization capture has been previously demonstrated to increase the analytical sensitivity of PCR-based detection of TB in pleural fluid (by 10- to 100-fold) (113) and in sputum (124, 125), and to increase the amount of background DNA tolerated for isothermal recombinase polymerase amplification (RPA) of HIV DNA in whole blood (109). Other forms of degraded DNA, such as formalin-fixed paraffin-embedded (FFPE) tissue or ancient DNA samples, may benefit from hybridization's ability to recover fragmented DNA. Hybridization is also compatible with the addition of chaotropic salts, such as guanidinium thiocyanate, which can simultaneously lyse cells and viruses and preserve fragile nucleic acids and may be particularly useful for capture of RNA (106, 151).

I have tested our sequence-specific capture method across a range of conditions but have not yet tested it in sample types other than urine. I have also not yet applied it to capture of longer DNA (>150 bp) or RNA, although we have begun early work to apply it to capture of HIV RNA. The results of previous studies (Table 2.2), however, suggest that hybridization is tolerant to a wide variety of complex sample matrices, supporting the possibility that our protocol may improve performance in other sample types. I have verified robust performance for DNA fragments from 25–150 bp, target concentrations from 0.5–10^4 copies/mL, and non-target DNA background from 0–10 µg.

One area where our protocol may require modification is for sample types with high protein concentrations, such as plasma or sputum. If the magnetic beads are added prior to denaturation, according to the current protocol, it is likely that as proteins denature, they will adsorb to the bead surface and impact bead behavior. I have observed that beads become more likely to aggregate after heating in urine samples with proteinuria or blood present, resulting in less effective collection of beads on the magnet during wash steps and a tendency for beads to transfer to qPCR during the elution step. The addition of a Proteinase K digestion step appears promising for mitigating bead loss in this situation (see Section 7.6). Another limitation resulting from the sequence-specific nature of hybridization is that it is designed for isolation and detection of specific target sequences, which will be most relevant for pathogen detection. Our sequence-specific capture method cannot be used to purify total DNA and is thus not appropriate for next-generation sequencing applications. Sequence-specific capture can, however, be multiplexed and ongoing work in our lab aims to develop a multiplexed system by adding

capture probes to target multiple genomic regions. I have demonstrated feasibility of including multiple capture probes, showing capacity for up to 10 probes with only a minor reduction in recovery (see Figure 7.3). Additional optimization specifically for multiplexed capture may enable higher-scale multiplexing.

3.10. CONCLUSIONS

To increase the diagnostic sensitivity of TB cfDNA detection in urine, we aimed to develop an improved sample preparation method for short DNA fragments. We selected sequence-specific purification as a DNA extraction technique likely to perform well for short, dilute cfDNA. To enable successful detection of low concentrations of TB-specific cfDNA in clinical urine samples, we pushed the limits of sequence-specific capture to achieve maximum analytical sensitivity and accommodate large-volume samples. Our final protocol improves upon both alternate urine cfDNA extraction methods (127) and previous sequence-specific capture protocols for sample preparation of targets other than urine cfDNA (Table 2.2). It reliably detects 5 copies of dsDNA spiked into 10 mL urine (0.5 copies/mL), recovers nearly 100% of short fragments independent of fragment length (25–150 bp), and enables amplification of cfDNA from variable 10 mL urine samples in a single qPCR well (500X concentration). Combining sequence-specific purification with an ultrashort qPCR design, we demonstrate detection of TB-specific DNA fragments as short as 25 bp with high efficiency. Our method meets several design criteria for short urine cfDNA not satisfied by previous urine cfDNA extraction methods and sequence-specific capture protocols: (1) high recovery of short fragments, (2) large input sample volume, and (3) <1 copy/mL sensitivity. Two key design features ensure its robust performance and unprecedented sensitivity: a two-probe system for recovery of both strands of dsDNA and dual biotinylated capture probes to facilitate optimal probe density, increase thermostability of the bead-probe linkage, and eliminate interference from endogenous biotin. The strong analytical performance of our sequence-specific capture method described here suggests that the approach will increase recovery of short, dilute urine cfDNA and contribute to increased sensitivity of TB cfDNA detection in urine. Design of capture probes for new targets is straightforward, and, in our experience, does not require design iterations or assay optimization to achieve high recovery. We believe our hybridization method will be versatile and useful for broad applications in the urine cfDNA field and may also have utility for other targets and sample types. We hope that the detailed, user-ready protocol and probe design guidelines included here will enable reproduction of our results, transfer of the assay to new users, and improved sample preparation for wide-ranging applications benefiting from sequence-specific purification.

3.11. Funding acknowledgements

Research reported in this chapter was supported by the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R21AI125975 and the National Science Foundation Graduate Research Fellowship Program.

4. Analytical comparison of urine cell-free DNA extraction methods

This chapter is adapted from an article published in Journal of Molecular Diagnostics, 21(6): Oreskovic A, Brault ND, Panpradist N, Lai JJ, Lutz BR. Analytical comparison of methods for extraction of short cell-free DNA from urine. Pg 1067-78. Copyright Elsevier (2019). https://doi.org/10.1016/j.jmoldx.2019.07.002.

This article was written to target a broader audience interested in urine cfDNA for applications beyond TB. Here, the data from this article have been updated to include the 10 mL surface-based hybridization capture method, which had not yet been developed at the time of publication.

4.1. Abstract

Urine cell-free DNA (cfDNA) is a valuable noninvasive biomarker for cancer mutation detection, infectious disease diagnosis (e.g., TB), organ transplantation monitoring, and prenatal screening. Conventional silica DNA extraction does not efficiently capture urine cfDNA, which is dilute (ng/mL) and highly fragmented (30–100 bp). The clinical sensitivity of urine cfDNA detection increases with decreasing target length, motivating use of sample preparation methods designed for short fragments. We compared the analytical performance of two published protocols (Wizard/GuSCN, Q Sepharose), three commercial kits (Norgen, QIAamp, MagMAX), and two in-house sequence-specific hybridization capture techniques (solution-based and surface-based). We characterized dependence on fragment length (25–150 bp), performance at low concentrations (10 copies/mL), tolerance to variable urine conditions, and susceptibility to PCR inhibition. Hybridization capture and Q Sepharose performed best overall (60–90% recovery), although Q Sepharose had reduced recovery (<10%) of the shortest 25 bp fragment. Wizard/GuSCN recovery was dependent on pH and background DNA concentration and was limited to <35% even under optimal conditions. The Norgen kit suffered from consistent PCR inhibition but had high recovery of short fragments. The QIAamp and MagMAX kits had minimal recovery of fragments <150 bp and <80 bp, respectively. Our results reveal that urine cfDNA extraction methods

differ widely in ability to capture short, dilute cfDNA in urine and suggest that employing sub-optimal methods may profoundly impair clinical results.

4.2. INTRODUCTION

Urine cell-free DNA (cfDNA) is an emerging noninvasive biomarker for cancer mutation detection (3, 38, 39, 126), infectious disease diagnosis (4–6, 52), organ transplant monitoring (43, 44), and prenatal screening (3, 45, 79). As cells die throughout the body, cfDNA is released into the bloodstream. A fraction of circulating cfDNA, composed largely of short fragments, crosses the kidney barrier, is excreted in urine, and can be analyzed by PCR or sequencing (3). This subset of urine cfDNA derived from circulating cfDNA is known as transrenal DNA (trDNA), but cfDNA can also be generated directly in urine from cells shed along the urinary tract.

To maximize the clinical sensitivity and reproducibility of urine cfDNA analysis, extraction methods capable of efficiently capturing short, dilute DNA fragments are essential. While plasma cfDNA is primarily nucleosomal with a peak length of 160–167 bp (46, 49, 50), urine cfDNA is more fragmented. The upper length limit of the transrenal fraction of urine cfDNA is defined by glomerular filtration, and all urine cfDNA fragments are quickly degraded further in urine (53). Determining the true length distribution of urine cfDNA is challenging because both extraction and library preparation methods may underestimate the presence of shorter fragments, but the majority of fragments are expected to be under 100 bp (45, 46, 55). Peak fragment length varies across patients, but may be as low as 30–60 bp (45, 46, 55).

In this chapter, I present an analytical comparison of two published urine cfDNA extraction protocols (Wizard/GuSCN and Q Sepharose), three commercial kits (Norgen, QIAamp, and MagMAX), and two versions of an in-house hybridization capture protocol (surface-based and solution-based). Due to the extensive fragmentation of urine cfDNA, the diagnostic clinical sensitivity of urine cfDNA detection increases with decreasing target length. Maximizing sensitivity by targeting shorter fragments is especially critical since urine cfDNA is also dilute, with total concentrations ranging from <1 ng/mL to 200 ng/mL (3, 47, 48) and copy numbers of specific targets much lower. In a study detecting fetal cfDNA in maternal urine, decreasing PCR amplicon length from 65 bp to 39 bp increased clinical sensitivity from 25% to 75%. A further decrease to 25 bp was required before achieving 100% detection (79). This effect may be even more pronounced for bacterial, viral, and mitochondrial cfDNA, which are not protected by histones and are therefore more degraded than human genomic cfDNA (49, 53, 55). For TB urine cfDNA, a modest 10 bp decrease in amplicon length (49 bp to 39 bp) led to 5- to 10-fold improvement in

detected concentration (81). Critically, the ability to target shorter cfDNA fragments lies not only in decreasing amplicon length, but also in design and selection of sample preparation methods capable of capturing and concentrating the short, dilute fragments that constitute the bulk of urine cfDNA.

Unfortunately, conventional extraction methods for cell-associated DNA or even plasma cfDNA are not suitable for urine cfDNA since they are not designed for short fragments. The “Boom” method, commonly used for both research and clinical work, adsorbs DNA to silica under chaotropic conditions (73). The key driving forces of silica adsorption are hydrophobic interactions due to dehydration of silica and DNA surfaces and hydrogen bonding between silica and the DNA backbone, both of which depend on DNA length (74). Consequently, silica adsorption is less effective at purifying short fragments, with recovery generally decreasing below 50–100 bp. Silica adsorption also requires relatively high DNA concentrations for optimal performance since a fraction of DNA may remain irretrievably bound to the silica surface (75, 76). This loss is trivial in most samples, but for low concentration samples like urine cfDNA it may make up a significant portion of the input.

With these limitations in mind, an ideal urine cfDNA extraction method would enable high recovery of short DNA from dilute solutions. Despite the great clinical promise of urine cfDNA as an easy-to-access sample, there has been little quantitative comparison of approaches taken to improve recovery of urine cfDNA. A recent review emphasized the lack of standardization in sample preparation methodology, including DNA extraction, as a key limitation in the development of urine cfDNA assays for TB (42). Previous studies have compared clinical detection rates (79, 81) and total cfDNA recovery (48, 91, 93) of a limited set of extraction methods, but none have investigated analytical performance using spiked samples. Other studies have compared additional pre-analytical variables (e.g., collection method, storage conditions, centrifugation speed) but did not include the extraction method in their comparisons (88, 92). For each extraction method compared here, I characterized the dependence on DNA fragment length, performance at low DNA concentrations, tolerance to variable urine conditions, and susceptibility to PCR inhibition.

This work was intended to go beyond the scope of TB and appeal to broader clinical audiences interested in urine cfDNA. The aim of this work was to gain a better understanding of the current landscape of urine cfDNA extraction methods and develop a representative dataset to help guide selection and optimization of DNA extraction methods for urine cfDNA analysis. Careful design of sample preparation methods should lead to increased clinical sensitivity and reproducibility of urine cfDNA diagnostics. The strong performance of hybridization capture in relation to other urine cfDNA extraction

methods also motivated continued development of an assay using this approach for diagnosis of TB from urine cfDNA.

4.3. Overview of urine cell-free DNA extraction methods

The Wizard Resin/Guanidinium Thiocyanate (Wizard/GuSCN) method uses high concentrations (>3M) of chaotropic guanidinium thiocyanate (GuSCN) to adsorb DNA to Wizard silica resin. This approach was used to originally demonstrate the presence of cfDNA in urine (3) and has since been widely applied, most frequently for detecting TB (4) and fetal (45) cfDNA. The Q Sepharose method uses a quaternary ammonium anion exchange resin to pre-concentrate DNA prior to desalting on a silica spin column. It improves recovery of short urine cfDNA fragments compared to Wizard/GuSCN (79) and has often been used to detect tumor cfDNA mutations for cancer diagnosis, monitoring, and prognosis (38, 39). The Norgen Biotek (Ontario, Canada) Urine Cell-Free Circulating DNA Purification Kit uses a hybrid silica/silicon carbide spin column, where addition of silicon carbide reportedly improves yield of short DNA compared to silica alone (US Patent 9,422,596). The Qiagen (Hilden, Germany) QIAamp Circulating Nucleic Acid Kit uses a silica vacuum column and reportedly improves recovery of fragmented DNA compared to other Qiagen kits. It is one of the most widely used commercial kits for plasma cfDNA extraction (152) but is not commonly used for urine cfDNA. The Thermo Fisher Scientific (Waltham, MA) MagMAX Cell-Free DNA Isolation Kit uses Dynabeads MyOne Silane to maximize binding kinetics and capacity but is intended primarily for plasma cfDNA. It was included as a reference method to represent best-case silica adsorption without modifications specifically for urine cfDNA, although it has been used previously in urine (56, 153). The two hybridization capture approaches developed in house and described in Chapter 3 (surface-based hybridization and solution-based hybridization) were also included in the comparison.

4.4. Methods

Synthetic DNA target design. To study the analytical performance of the urine cfDNA extraction methods, I spiked synthetic single-stranded DNA (ssDNA) targets into pooled urine prior to extraction and analysis by quantitative PCR (qPCR). The targets were selected from a conserved and specific region of the insertion sequence IS6110 of the *Mycobacterium tuberculosis* complex (GenBank Accession #X17348) (130). The targets were 25, 40, 80, and 150 bp in length, as listed in Table 4.1. The 40, 80, and 150 bp targets were designed to be amplified by a shared primer set, with additional bases outside of the primer amplification region added to the 3' end of the 40 bp target to create the 80 bp and 150 bp

targets. The 25 bp target was designed to be amplified by a separate set of primers in a two-stage, single tube PCR for ultrashort targets (79).

Table 4.1: Probe, primer, and target sequences. *Single-stranded synthetic oligonucleotides were used as spike-in targets to study the analytical performance of urine cfDNA extraction methods. Shared sequence between the 40, 80, and 150 bp targets is bolded. Binding regions for the biotinylated capture probes are underlined.*

Assay	Oligonucleotide	Sequence
PCR of 40, 80, 150 bp targets	Forward primer	5' – CGAACCCTGCCCAGGTCGA – 3'
	Reverse primer	5' – GTAGCAGACCTCACCTATGTGT – 3'
	40 bp target	5' – **CGAACCCTGCCCAGGTCGACACATAGGTGAGGTCTGCTAC** – 3'
	80 bp target	5' – **CGAACCCTGCCCAGGTCGACACATAGGTGAGGTCTGCTAC**ACACCATTCA ATTTCATCACTGCCAATACTCCACTCTCAT – 3'
	150 bp target	5' – **CGAACCCTGCCCAGGTCGACACATAGGTGAGGTCTGCTAC**ACACCATTCA ATTTCATCACTGCCAATACTCCACTCTCATCTACACAACCCATTAGTACCTTA CCTCGCTTCCTATCCCAATTCACTTAATCTTAAACCGGTCAGGGAAG – 3'
PCR of 25 bp target	25 bp target	5' – CCGGCTGTGGGTAGCAGACCTCACC – 3'
	1st stage hairpin forward primer	5' – GCGTAAGAAT/iMe-isodC/AAACGTCGCTCAACTTCCATT CTTACGCCCGGCTGTGG – 3'
	2nd stage universal forward primer	5' – AACGTCGCTCAACTTCCATT – 3'
	Reverse primer	5' – TTAGAGAAGGTGAGGTCTGC – 3'
	MGB TaqMan probe	5' – 6FAM/CCGGCTGTGGGTA/MGBNFQ – 3'
Surface-based hybridization capture (10 mL)	Biotinylated capture probe for 40, 80, 150 bp targets	5' - /52-Bio/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/ - 3'
	Biotinylated capture probe for 25 bp target	5' - /52-Bio/AAAAAAAAAAAAAAAAAAAAGGTGAGGTCTGCTAC/3SpC3/ - 3'
Solution-based hybridization capture (1 mL)	Biotinylated capture probe for 40, 80, 150 bp targets	5' – /5BiosG/AGACCTCACCTATGTGTC/3SpC3/ – 3'
	Biotinylated capture probe for 25 bp target	5' – /5BiotinTEG/GAGGTCTGCTACCCA/3SpC3/ – 3'

All oligonucleotides were ordered from Integrated DNA Technologies (Coralville, IA), except for the MGB TaqMan probe, which was from Thermo Fisher Scientific (Waltham, MA).

DNA extraction from pooled human urine. Urine from five healthy volunteers was pooled into a representative sample, supplemented with 10 mM EDTA, and stored at -80°C until analysis.

Wizard resin/guanidinium thiocyanate: 5 mL urine was mixed with 7.5 mL 6 M GuSCN and 1 mL Wizard Minipreps DNA Purification Resin (Promega, Madison, WI), rotated at room temperature for 2 hours, and vacuum filtered through a syringe fitted with a Wizard minicolumn. The resin was washed twice with 5 mL wash buffer (80 mM KOAc, 8.3 mM Tris-HCl pH 7.5, 40 µM EDTA, 55% ethanol). The

minicolumn was removed and dried (10,000g, 2 min). DNA was eluted with 100 µL 60°C nuclease-free water (1 min incubation, 1 min 16,000g).

Q Sepharose anion exchange resin: 10 mL urine was mixed with 300 µL Q Sepharose Fast Flow (GE Healthcare, Waukesha, WI) and rotated at room temperature for 30 min. The resin was pelleted (1,800g, 5 min), resuspended in 1 mL low salt buffer (0.3 M LiCl, 10 mM NaOAc pH 5.5), transferred to a Mini Bio-Spin Column (Bio-Rad Laboratories, Hercules, CA), and filtered (800g, 1 min). The resin was washed with 4 x 0.5 mL low salt buffer (800g, 30s). DNA was eluted (800g, 3 min) using 670 µL high salt buffer (2 M LiCl, 10 mM NaOAc pH 5.5). The eluate was mixed with 2 mL 95% ethanol and applied incrementally to a QIAquick column (Qiagen) (800g, 30s). The column was washed twice with 0.5 mL 2 M LiCl in 70% ethanol and twice with 0.5 mL 75 mM KOAc pH 5.5 in 80% ethanol (800g, 30s). The column was dried (20,000g, 3 min) and DNA eluted (20,000g, 2 min) in 106 µL elution buffer (Qiagen).

Norgen Urine Cell-Free Circulating DNA Purification Mini Kit: DNA was extracted from 2 mL urine using the manufacturer's protocol and eluted into 50 µL.

Qiagen QIAamp Circulating Nucleic Acid Kit: DNA was extracted from 4 mL urine using the manufacturer's protocol for purification of circulating nucleic acids from urine and eluted into 50 µL. The QIAamp experiments were carried out later than those for other methods, so a different urine sample was used.

Thermo Fisher MagMAX Cell-Free DNA Isolation Kit: DNA was extracted from 1 mL urine using the manufacturer's protocol for manual isolation of cfDNA from urine and eluted into 20 µL.

Solution-based hybridization capture: 1 mL urine was mixed with 15 nM biotinylated capture probe (Table 4.1), 1 M NaCl, and 10 mM Tris-HCl pH 7.5, denatured (95°C, 10 min), and hybridized (45°C, 15 min). Hybridized complexes were immobilized on 83.2 µL Dynabeads MyOne Streptavidin C1 (Thermo Fisher) by 15 min rotation at room temperature. Beads were washed twice with 1 mL high salt wash (1 M NaCl, 10 mM Tris-HCl pH 7.5) and once with 1 mL low salt wash (15 mM NaCl, 10 mM Tris-HCl pH 7.5). DNA was eluted (20 µL 20 mM NaOH) and partially neutralized (3.5 µL 100 mM HCl). A detailed protocol is also given in Appendix A, Section 9.2.1.

Surface-based hybridization capture: 50 µL Dynabeads MyOne Streptavidin C1 (Thermo Fisher) were washed three times with an equal volume bead wash buffer (1M NaCl, 10 mM Tris-HCl pH 7.5, 0.05% Tween-20). 50 pmol of biotinylated capture probe (Table 4.1) was immobilized on the beads by 15 min rotation at room temperature. Note that for the experiments in this chapter, only a single biotinylated

probe was used (to target ssDNA) rather than two probes used in the final protocol (to capture both strands of dsDNA). Unbound streptavidin was blocked by the addition of 12.5 nmol free D-biotin (VWR) and 15 min rotation at room temperature. The beads were again washed three times with an equal volume bead wash buffer (1M NaCl, 10 mM Tris-HCl pH 7.5, 0.05% Tween-20). 10 mL urine was mixed with 50 µL functionalized beads, 1 M NaCl, 10 mM Tris-HCl pH 7.5, and 0.05% Tween-20, denatured (>90°C, 15 min), and hybridized (room temperature, 30 min). The beads were pelleted (5000g, 5 min) and all but 1 mL of supernatant was discarded. The beads were resuspended in the remaining supernatant and washed twice with 1 mL high salt wash (1 M NaCl, 10 mM Tris-HCl pH 7.5, 0.05% Tween-20) and once with 1 mL low salt wash (15 mM NaCl, 10 mM Tris-HCl pH 7.5). DNA was eluted (20 µL 20 mM NaOH) and partially neutralized (3.5 µL 100 mM HCl).

Quantitative PCR. qPCR of experimental samples was performed in triplicate in a 50 µL volume containing 5 µL of DNA output except for hybridization capture experiments, where the entire output (~23 µL) was analyzed in a single PCR well. To control for contamination, no template controls (NTCs) were run not only for PCR (n=3) but also through the entire DNA extraction procedure for all experiments ($n \geq 3$).

qPCR of the 40, 80, and 150 bp targets was carried out in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with an initial incubation of 94°C for 5 min followed by 45 amplification cycles (94°C for 30s, 58°C for 30s, and 68°C for 1 min). Each reaction contained 1.25 U OneTaq Hot Start DNA Polymerase (New England Biolabs, Ipswich, MA), 1X OneTaq GC Reaction Buffer (NEB; 80 mM Tris-SO_4, 20 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 5% glycerol, 5% DMSO, 0.06% IGEPAL CA-630, 0.05% Tween 20, pH 9.2), 0.8 mM dNTPs (NEB), 0.4X EvaGreen (Biotium, Fremont, CA), 200 nM forward primer, and 200 nM reverse primer (Table 1). C_q values were determined using the CFX Manager Software Version 3.1 at a threshold of 500 RFU and recovered copies were calculated by a standard curve. Note that at the time of this study, I had not yet improved PCR specificity using LNA-substituted primers, so an earlier version of the PCR primers and cycling protocol was used.

Ultrashort qPCR of the 25 bp target was carried out in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with an initial denaturation phase (94°C for 5 min), 10 pre-amplification cycles to extend the first stage loop primer (94°C for 30s and 45°C for 1 min), and 40 amplification cycles (94°C for 30s and 59°C for 1 min). Each reaction contained 1.25 U Hot Start Taq DNA Polymerase (NEB), 1X Standard Taq Buffer (NEB; 10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3) supplemented with an additional 0.5 mM $MgCl_2$ and 70 mM Tris-HCl, 0.8 mM dNTPs (NEB), 50 nM first stage hairpin forward primer, 700 nM

second stage universal forward primer, 700 nM reverse primer, and 100 nM MGB TaqMan probe (Table 1). Quantification cycle (C_q) values were determined using the CFX Manager Software Version 3.1 at a threshold of 100 RFU and recovered copies were calculated by a standard curve.

4.5. RESULTS OF ANALYTICAL CHARACTERIZATION

Table 4.2 summarizes the urine cfDNA extraction methods, including processing time, cost, and volume of urine analyzed.

Table 4.2: Overview of urine cfDNA extraction methods.

Urine cfDNA extraction method	Key purification chemistry	Rationale for selection	Processing time (hands on time)[a]	Cost per sample[b]	Effective urine volume analyzed per PCR well
Surface-based Hybridization Capture (10 mL)	Hybridization to biotinylated probe pre-immobilized on streptavidin magnetic beads	Developed in our lab specifically for urine cfDNA; demonstrated high recovery of short cfDNA	1.75 hr (1 hr)	$9	10000 µL[c]
Solution-based Hybridization Capture (1 mL)	Hybridization to biotinylated probe followed by capture on streptavidin magnetic beads	Developed in our lab specifically for urine cfDNA; demonstrated high recovery of short cfDNA	1.75 hr (1 hr)	$15	1000 µL[c]
Wizard/GuSCN	Adsorption to silica resin in presence of high concentration chaotrope (3 – 6M)	Originally used to isolate cfDNA from urine; widely used in the literature	3 hr (2.5 hr)	$5	472 µL
Q Sepharose	Pre-concentration by anion exchange resin followed by adsorption to silica spin column	Shown to improve recovery of short fragments compared to Wizard/GuSCN method	3 hr (1 hr)	$5	250 µL
Norgen Urine Cell-Free Circulating DNA Purification Kit	Adsorption to silica/silicon carbide hybrid spin column in presence of chaotrope	Commercial kit designed specifically for urine cfDNA	1.5 hr (1.25 hr)	$5	200 µL
QIAGEN QIAamp Circulating Nucleic Acid Kit	Adsorption to silica vacuum column in presence of chaotrope	Commercial kit commonly used for plasma cfDNA	2 hr (1.25 hr)	$25	400 µL
Thermo Fisher MagMAX Cell-Free DNA Isolation Kit	Adsorption to Dynabeads MyOne Silane in presence of chaotrope	Representative commercial silica kit; "best case" scenario without specific designs for urine cfDNA	2.5 hr (2 hr)	$18	250 µL

[a] For 12 samples; sample preparation time only, not including qPCR.
[b] Sample preparation cost only, not including qPCR.
[c] All qPCR carried out using 5 µL of sample per well expect for hybridization, where the entire ~23 µL output was analyzed in a single PCR well.

4.5.1. *Effect of DNA fragment length on recovery*

To evaluate the dependence of urine cfDNA extraction methods on fragment length, I extracted DNA from urine spiked with 10^4 copies/mL of synthetic ssDNA target of length 25, 40, 80, or 150 bp (Figure 4.1A). Figure 4.1B shows the percent recovery of each extraction method across fragment lengths. Hybridization capture was the only method that maintained high recovery (90–110% and 73–84% for surface-based and solution-based hybridization, respectively) across all fragment lengths from 25–150 bp. Q Sepharose also had high recovery (63–75%) of 40–150 bp fragments, but reduced recovery (9%) of the shortest 25 bp fragment. Wizard/GuSCN recovery was initially very low (<5%) across all fragments. Later experiments showed that Wizard/GuSCN was dependent on urine composition, particularly pH and background DNA. Even after adjusting urine to optimal conditions (pH 6, 1000 ng/mL sheared salmon sperm DNA, Thermo Fisher), Wizard/GuSCN recovery was still low (9–17%) across 40–150 bp fragments and further reduced (2%) for the 25 bp fragment. The Norgen kit had moderate recovery (30–41%) across 40–150 bp fragments and improved recovery (72%) of the 25 bp fragment. Recovery using the QIAamp kit was limited (18%) for the longest 150 bp fragment and was very low (1% and 0.2%, respectively) for the shorter 80 bp and 40 bp fragments. The MagMAX kit recovery was high (66%) for the 150 bp fragment, but quickly diminished with decreasing fragment length and was practically nonexistent (0.2%) for the 40 bp fragment.

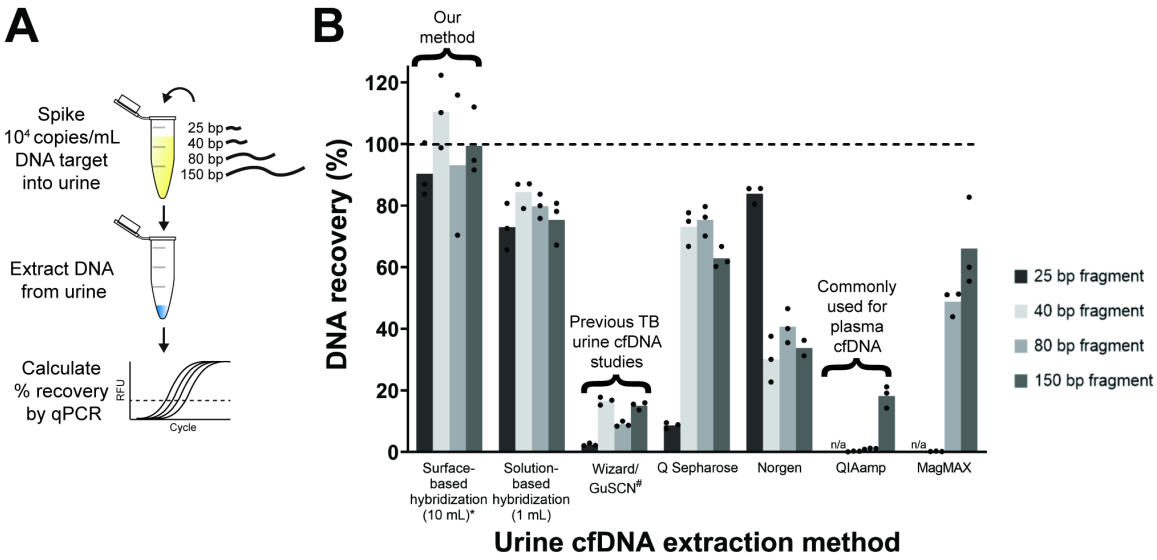


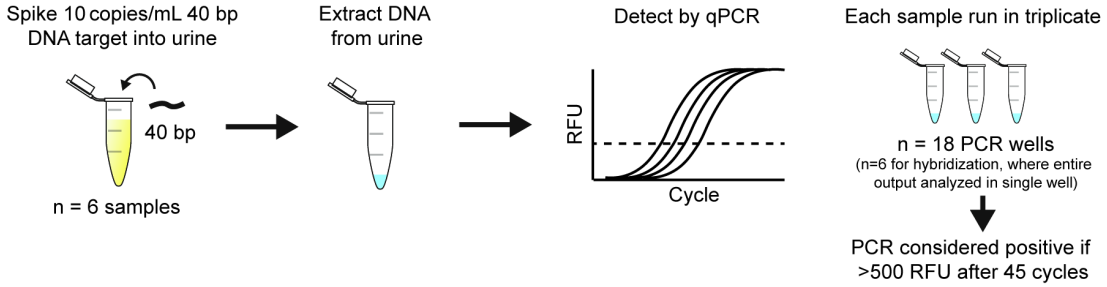
Figure 4.1: Hybridization is the only urine cfDNA extraction method that maintains high recovery across all DNA fragment lengths. (A) *To evaluate the length dependence of urine cfDNA extraction methods, synthetic ssDNA targets of various lengths (25, 40, 80, and 150 bp) were spiked into urine at 10^4 copies/mL prior to extraction.* ***(B)*** *Mean percent recoveries for DNA spiked into urine are shown for each extraction method (n=3 technical replicates processed on the same day). *Surface-based*

hybridization was tested at 100 copies/mL instead of 10^4 copies/mL, and the experiment was repeated three times across different days. For this extraction method only, the data points represent the mean recovery of n=3 technical replicates from each independent experiment. #Wizard/GuSCN samples were adjusted to optimal pH 6 and spiked with 1000 ng/mL background DNA prior to extraction.

4.5.2. *Ability to detect low concentrations of short DNA fragments*

To determine each method's potential for sensitive capture of short, dilute urine cfDNA, I spiked 10 copies/mL of 40 bp target into urine before extraction (Figure 4.2A). As shown in Figure 4.2B, hybridization capture (surface-based and solution-based) and Q Sepharose reliably yielded detectable DNA from all low concentration spiked samples. Surface-based hybridization capture was able to detect even lower concentrations down to at least 1 copy/mL. The Norgen kit also detected all samples, but only weakly. Wizard/GuSCN weakly detected 83% of samples but was inconsistent, with only 44% positive PCR wells. The QIAamp and MagMAX kits did not allow confident detection of any positive samples.

A



B

Urine cfDNA extraction method	Samples with ≥1 positive PCR well	Positive PCR wells	Detected copies/well (mean ± SD)	Expected copies/well (if 100% recovery)
Surface-based hybridization (10 mL)	100% (6/6)	100% (6/6)	14.4 ± 5.8*	10*
Solution-based hybridization (1 mL)	100% (6/6)	100% (6/6)	7.8 ± 5.1	10
Wizard/GuSCN	83% (5/6)	44% (8/18)	1.6 ± 0.87	2.5
Q Sepharose	100% (6/6)	89% (16/18)	5.0 ± 1.5	4.7
Norgen	100% (6/6)	89% (16/18)	1.42 ± 0.67	2
QIAamp	17% (1/6)	6% (1/18)	0.10 ± 0.17	4
MagMAX	33% (2/6)	11% (2/18)	0.77 ± 0.31	2.5

***Figure 4.2: Hybridization, Q Sepharose, and Norgen methods can detect low concentrations of short cfDNA fragments. Wizard/GuSCN, QIAamp, and MagMAX methods are not expected to perform well under these conditions**. **(A)** 10 copies/mL of 40 bp ssDNA target was spiked into urine, extracted, and detected by qPCR (n=6 technical replicates processed on the same day). PCR was carried out in triplicate, except for hybridization, where the entire output was analyzed in a single PCR well. **(B)** The ability of each urine cfDNA extraction method to detect low concentration samples. Expected copies/well for theoretical 100% recovery was calculated based on the initial 10 copies/mL target concentration and adjusted for the urine input and elution volume of each method. *Surface-based hybridization was tested at 1 copy/mL instead of 10 copies/mL.*

4.5.3. *Tolerance to varied urine conditions*

To test the methods' tolerance to varied conditions expected in urine, I extracted 10^4 copies/mL of 150 bp target from buffer (PBS or TBS) with a range of pH (pH 5, 6, 7, 8), background DNA (0, 100, 1000 ng/mL sheared salmon sperm DNA, Thermo Fisher), and salt (13.7, 137, 500 mM NaCl) conditions. I found that the Wizard/GuSCN method was highly dependent on urine composition, specifically pH and background DNA. Recovery decreased as pH increased above pH 6 (Figure 4.3A). Spiking in background DNA (1 µg/mL) improved recovery, but maximum recovery was still well below that of other methods (Figure 4.3B). Variation in salt from 13.7–500 mM NaCl had no effect on recovery (Appendix C, Table 11.1). The remaining methods were all relatively tolerant to variations in pH, background DNA, and salt (Appendix C, Table 11.1). Recovery by the Q Sepharose method was moderately reduced with very low background DNA (≤10 ng/mL) but was consistent across biological urine replicates (Figure 4.4).

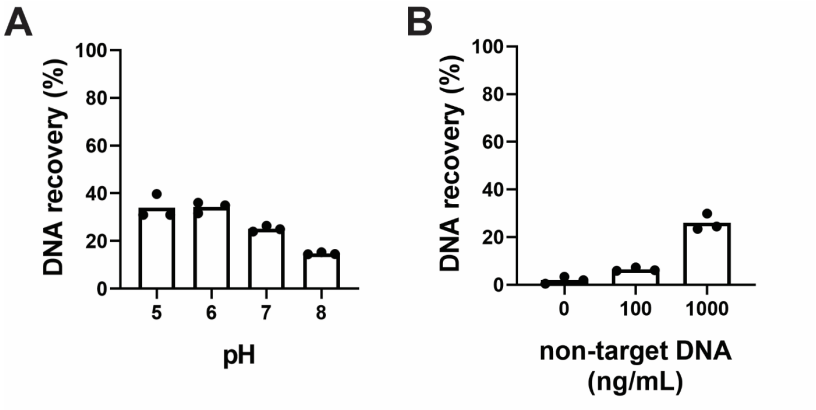


Figure 4.3: The Wizard/GuSCN method is dependent on urine pH and background DNA concentration. ***(A)*** *Recovery decreases with increasing pH. 10^4 copies/mL of 150 bp ssDNA target were spiked into PBS with 1000 ng/mL background DNA (mean of n=3 technical replicates processed on the same day).* ***(B)*** *Recovery increases with the addition of background DNA. 10^4 copies/mL of 150 bp ssDNA target were spiked into PBS pH 6 (mean of n=3 technical replicates processed on the same day).*

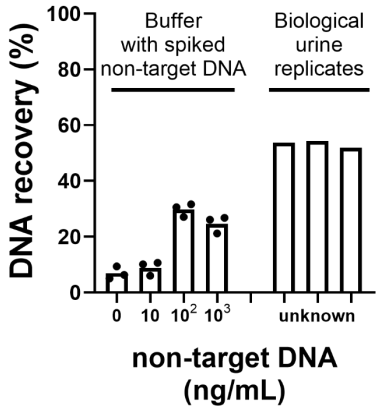


Figure 4.4: Dependence of Q Sepharose on background DNA concentration in buffer and performance across biological urine replicates. *10^4 copies/mL of 150 bp ssDNA target were spiked into 1X TBS (n=3 technical replicates processed on the same day) or urine (n=3 biological replicates processed on the same day) prior to extraction. In buffer, the Q Sepharose method had lower recovery at lower background DNA concentrations (0, 10 ng/µL) than at higher background DNA concentrations (100, 1000 ng/µL). Unlike the Wizard/GuSCN method, however, the 100 ng/mL threshold is within the expected clinical range of urine cfDNA concentrations likely to be found in patient samples. In urine, recovery was consistent across biological urine replicates despite the moderate dependence on background DNA concentration in buffer. All urine samples had higher recovery than seen in buffer. Based on these results, I classified Q Sepharose as tolerant to variable urine conditions, but there is a possibility that unusually dilute samples may have reduced recovery.*

4.5.4. *Susceptibility to PCR inhibition*

To test for PCR inhibition resulting from each method, I extracted cfDNA from negative control urine, without added target, and spiked the resulting eluate (0, 1, 5, 10, or 20 µL) into 50 µL PCR containing 1000 copies of 40 bp target (Figure 4.5A). PCR inhibition was indicated by an increase in C_q. Hybridization capture (surface-based and solution-based), Wizard/GuSCN, Q Sepharose, and QIAamp were resistant to inhibition for up to 40% eluate (Figure 4.5B). While not accompanied by an increase in C_q, increasing the fraction of Wizard/GuSCN and Q Sepharose eluate reduced plateau RFU (Figure 4.5C). MagMAX suffered from slight inhibition at 20% eluate and severe inhibition at 40% eluate. Norgen led to inhibition at all conditions tested and no amplification at 40% eluate.

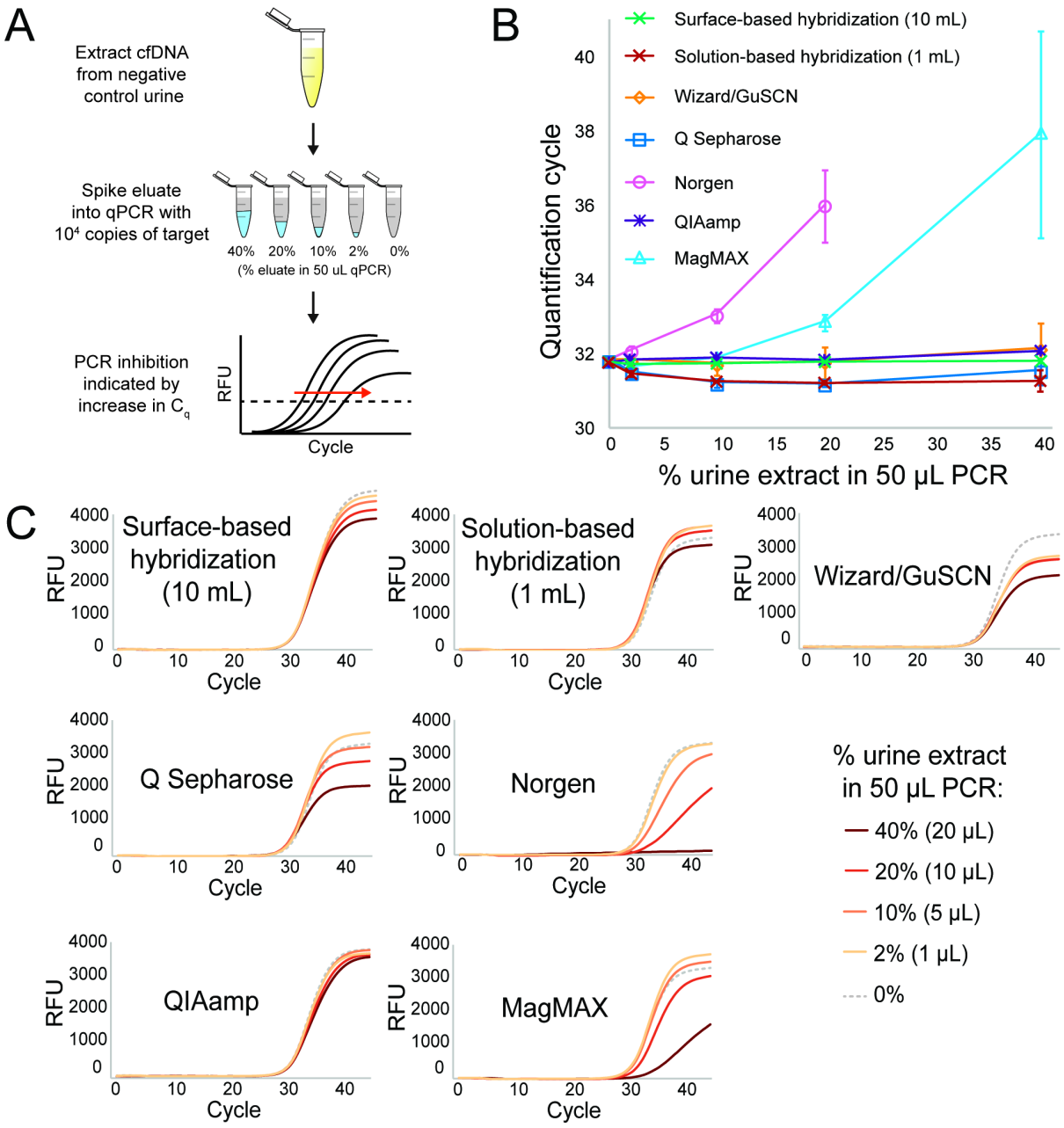


Figure 4.5: Urine cfDNA extraction methods have varying susceptibility to cause PCR inhibition. (A) *Eluate extracted from negative control urine (no added target) was spiked into PCR containing a constant target concentration (0, 1, 5, 10, or 20 µL eluate in 50 µL PCR, for final 0%, 2%, 10%, 20%, or 40% eluate, respectively). An increase in quantification cycle* ***(B)*** *indicates PCR inhibition (mean±SD, n=3 technical replicates processed on the same day). Representative PCR curves are shown in* ***(C).***

4.6. Discussion of each method's suitability for short, dilute urine cell-free DNA

The highly fragmented and dilute nature of urine cfDNA (30–100 bp or less, <1–200 ng/mL) (3, 45–48, 55) motivates the use of sample preparation methods capable of recovering short DNA with high efficiency. My goal was to generate a representative dataset to aid in selection and optimization of extraction methods to ensure high quality results from urine cfDNA studies. I characterized the analytical dependence of seven methods on a key set of variables to gain insight into how the methods may perform in clinical samples and to identify any critical pitfalls. In the subsections below, I discuss the strengths and limitations of each method and present final conclusions regarding choice of urine cfDNA extraction method. Apart from hybridization capture, which I developed specifically for detection of TB urine cfDNA, I strove to follow published protocols as closely as possible. Further optimizations including adjusting urine pH, spiking in background DNA, improving elution efficiency, and tailoring sample, elution, and PCR volumes may improve outcomes.

4.6.1. *Wizard resin/guanidinium thiocyanate*

Despite its use in the first study isolating urine cfDNA (3), the Wizard/GuSCN method has low and variable recovery. After observing low recovery (<5%) and significant variation across urine samples in preliminary work, I discovered that recovery was highly dependent on pH and background DNA concentration, both of which fluctuate widely across clinical urine samples. Urine pH ranges from 5–8 (mean 5.99–6.43) (131), but Wizard/GuSCN suffered at pH above 6. As silanol groups become deprotonated at higher pH, increased electrostatic repulsion between DNA and silica diminishes adsorption (74–76). Recovery by Wizard/GuSCN also improved as background DNA increased up to 1 µg/mL. I did not test concentrations >1 µg/mL, already well above the expected biological range of <1–200 ng/mL (3, 47, 48). Again, this limitation is not surprising for silica. Supplementation with carrier nucleic acids improves silica extraction yields, particularly for dilute samples, and is often implemented in commercial purification kits (154). Limited recovery of dilute DNA may be due to ineffective elution rather than inefficient adsorption. After adsorption in the presence of high concentration chaotrope, like in the Wizard/GuSCN method, a fraction of DNA may remain irretrievably bound due to strong hydrophobic interactions with silica (75, 76). The resulting DNA loss is particularly detrimental for dilute samples, where the irretrievable fraction represents a significant portion of the input (75).

The limited overall recovery and dependence of Wizard/GuSCN on urine composition, and thus its likely failure in a portion of patient samples, may partially explain the low, variable clinical sensitivities for TB diagnosis previously reported when using it as a sample preparation method for urine cfDNA. Notably,

the three most well-designed previous TB urine cfDNA studies (Cannas *et al.*, Labugger *et al.*, Patel *et al.*; see Table 2.1) all used the Wizard/GuSCN or similar silica resin-based methods (4–6). Even under ideal conditions, maximum recovery using the Wizard/GuSCN method was limited to <20% from urine and 30–35% from buffer. Although it resulted in improved recovery of moderately short targets (40 bp) compared to conventional silica adsorption (e.g., MagMAX), it was still unable to recover the shortest 25 bp fragment. Based on my analytical characterization, I do not recommend Wizard/GuSCN for use in clinical samples (particularly for low concentration targets) without further optimization. If used, I suggest adjusting urine samples to pH 5–6, spiking in ≥1 µg/mL carrier nucleic acid, and using elevated temperature and incubation time to increase elution yield (76).

4.6.2. *Q Sepharose anion exchange resin*

The Q Sepharose method improves upon Wizard/GuSCN in both recovery of short DNA fragments and overall yield. Q Sepharose had high recovery (63–75%) of fragments down to 40 bp. Previous comparison showed that Q Sepharose improved the clinical detection rate of fetal cfDNA in maternal urine compared to Wizard/GuSCN, and increased the detected concentration from 26.3 (range 0–46.9) to 196.5 (range 17.1–451.8) genome equivalents/mL urine (79). In the urine of a single TB patient, Q Sepharose doubled the concentration of detectable TB cfDNA compared to Wizard/GuSCN when using a short 39 bp PCR target. When using longer PCR targets (49 bp or 99 bp), the difference between Q Sepharose and Wizard/GuSCN disappeared, confirming the retention of shorter fragments by Q Sepharose (81). Q Sepharose has not been tested in a larger cohort of patients for TB diagnosis but would have the potential to improve upon studies that used Wizard/GuSCN.

My results support the conclusions of previous work and suggest that pre-concentration of urine cfDNA using anion exchange resin helps compensate for the length and concentration dependence of silica adsorption. Q Sepharose did not, however, completely overcome fragment length dependence, with <10% recovery of the shortest 25 bp fragment. In addition to concentrating cfDNA, Q Sepharose eliminates urine variabilities like pH that might otherwise affect silica adsorption. I expect Q Sepharose to perform well in clinical samples, as supported by its successful previous implementation for liquid biopsies (38, 39).

I recommend Q Sepharose as an established, ready-to-go protocol that would be sufficient for most applications. When extreme sensitivity and retention of the very shortest fragments are required, hybridization capture is likely the better option. Q Sepharose would be well-suited for next-generation sequencing, where total, not sequence-specific, purification of cfDNA is necessary. It should ideally be

paired with single-stranded library preparation, which has been shown to improve sequencing yield of <100 bp cfDNA fragments (49). For amplification applications, its resistance to PCR inhibition suggests that larger effective volumes could be amplified per reaction to increase analytical sensitivity. However, in my subsequent implementation of the Q Sepharose method for extraction of TB urine cfDNA prior to sequencing, I found that some clinical urine samples led to clogging of the silica columns and/or downstream PCR inhibition (see Section 6.7) that was not observed with healthy control urine samples.

4.6.3. *Norgen Urine Cell-Free Circulating DNA Purification Kit*

The Norgen kit had moderate recovery (30–41%) of fragments 40–150 bp, but higher recovery (72%) of the 25 bp fragment. It was the only silica-based method to efficiently capture the shortest fragment, demonstrating that hybrid silica/silicon carbide spin columns improve capture of ultrashort fragments relative to silica alone, as claimed by the manufacturer. Unfortunately, the Norgen kit also led to consistent PCR inhibition even when using only a small volume of eluate in PCR. Consequently, quantification using the Norgen kit is unreliable since each individual PCR assay will be uniquely affected by inhibition (155). As an example, I show the differential inhibition of the 40 bp and 25 bp qPCR assays in Figure 4.6. The Norgen kit was weakly capable of detecting low concentrations of DNA, but its analytical sensitivity is limited by a relatively small urine input (2 mL) combined with inhibition-restricted PCR volume. While not ideal for precise quantification or sensitive detection of dilute targets, the Norgen kit is a commercially available, user-friendly option. It could be used for quick, preliminary urine cfDNA analyses where qualitative or semi-quantitative detection is adequate.

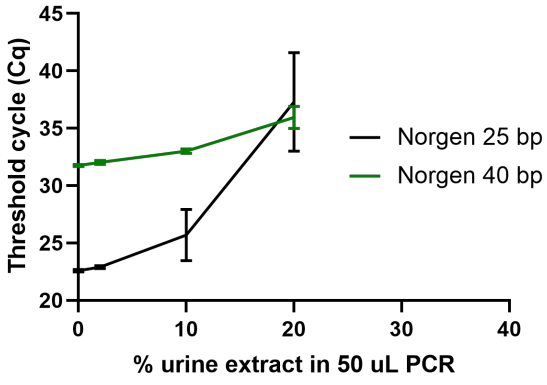


Figure 4.6: Differential PCR inhibition after extraction using Norgen kit. *10^4 copies of 40 bp or 25 bp ssDNA target were amplified in the presence of 0, 1, 5, 10, or 20 µL eluate from negative control urine extracted using the Norgen kit (0, 2, 10, 20, or 40% of 50 µL PCR, respectively). An increase in C_q indicates PCR inhibition (mean±SD, n=3 technical replicates processed on the same day). Here, amplification of the 25 bp target was more susceptible to PCR inhibition than amplification of the 40 bp target. Neither*

target amplified in the presence of 40% eluate. This result demonstrates how quantification of urine cfDNA after extraction using the Norgen kit would be unreliable due to differential PCR inhibition of different qPCR assays. Calculation of starting quantity based on C_q alone would be biased by the extent of PCR inhibition.

4.6.4. *Qiagen QIAamp Circulating Nucleic Acid Kit*

The QIAamp kit had limited recovery (18%), even for the longest 150 bp fragment. It also showed a clear dependence on fragment length, with significantly reduced recovery of the 80 bp and 40 bp fragments (1% and 0.2%, respectively). To confirm that the low observed recovery was not due to the urine sample used or errors in the extraction procedure, I also tested a long 400 bp double-stranded DNA (dsDNA) target, which had higher recovery (83% ± 4%, Figure 4.7). This result indicates that the low recovery using the QIAamp kit was due to the short length and/or single-stranded nature of the spiked target.

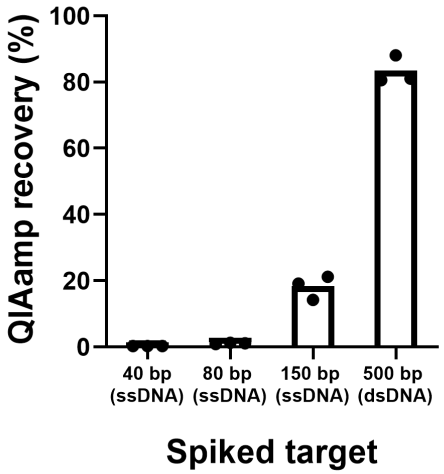


Figure 4.7: Recovery of QIAamp kit increases for longer double-stranded target. *10^4 copies/mL of 40 bp, 80 bp, 150 bp (ssDNA) or 400 bp (dsDNA) targets were spiked into urine and extracted using the Qiagen QIAamp Circulating Nucleic Acid Kit. Percent recovery is given as mean of n=3 extractions (technical replicates processed on the same day). While the recovery of short single-stranded fragments was low (<20%) using the QIAamp kit, recovery increased (83%) to match the kit's expected performance for a longer double-stranded target. This indicates that the low recovery of the original 40 bp, 80 bp, and 150 bp fragments was due to their short length and/or single-stranded nature, rather than a failure of the QIAamp kit protocol in our lab.*

The poor performance of the QIAamp kit for short fragments in urine is surprising given its widespread successful use in plasma. Several comparative studies have identified the QIAamp kit as one of the best-performing commercial options for plasma cfDNA (156–159). Although the kit has high overall yields from plasma, its recovery has been previously shown to suffer as fragment length decreases. For spiked dsDNA >100 bp, the QIAamp kit has >80% recovery from plasma (157, 159), but for dsDNA ≤100 bp the

recovery is reduced, with no recovery of a 25 bp fragment (159). This trend is in line with the manufacturer's product information, which claims efficient recovery of fragments down to 75 bp only. Regardless of sample type, both my results and others suggest that the QIAamp kit is inadequate for capturing short DNA fragments. It is unclear what caused the overall recovery from urine seen here to be lower than that of previous reports from plasma, but it may be at least partly due to the strandedness of the spiked target. Previous plasma studies used dsDNA (157, 159), while I used ssDNA, which interacts differently with silica surfaces on a molecular level (160). Although ssDNA has been reported to bind more strongly to silica at low pH than dsDNA (160), the relative recovery of ssDNA and dsDNA can be tuned by employing chaotropic binding buffers of different compositions (i.e., higher pH) (161). The specific buffer conditions of the QIAamp kit may be better suited for dsDNA than for ssDNA, exacerbating the existing length dependence when using a ssDNA target.

Due to the QIAamp kit's inefficient recovery of short fragments, particularly those that are single-stranded, I do not recommend its use in urine, where cfDNA is more fragmented than in plasma. Although the kit's performance may improve for dsDNA, an ideal urine cfDNA kit would be able to efficiently capture the full diversity of degraded urine cfDNA, which is likely to be a heterogenous mixture of short ssDNA, dsDNA, and damaged or nicked DNA.

4.6.5. *Thermo Fisher MagMAX Cell-Free DNA Isolation Kit*

The MagMAX kit was extremely dependent on fragment length, as expected for a silica-based method. It had high recovery of longer fragments but no detectable recovery of the 40 bp fragment. I do not recommend its use in urine samples, where the majority of cfDNA fragments are too short to be recovered efficiently. Other silica-based plasma cfDNA extraction kits may suffer from length-based limitations like the QIAamp and MagMAX kits. I caution against using plasma cfDNA kits for urine cfDNA extraction without experimentally verifying their ability to capture short DNA fragments.

4.6.6. *Hybridization capture*

Our lab identified hybridization capture as a sample preparation method likely to perform well for short, dilute urine cfDNA. Unlike silica adsorption, hybridization should be agnostic to both fragment length and concentration and robust against variations in clinical urine samples. My results confirmed that hybridization capture was the only method to maintain high recovery (90–110% and 73–84% for surface-based and solution-based hybridization, respectively) across all fragment lengths tested, even down to the shortest 25 bp fragment (90% and 73%, respectively). Hybridization capture was capable of reliably detecting low DNA concentrations (down to at least 1 copy/mL for surface-based hybridization

capture) and was tolerant to changes in urine pH, salt, and background DNA, suggesting that it will be effective in clinical samples. The small elution volume (20 µL) and complete removal of PCR inhibition enable the entire output from up to 10 mL urine to be analyzed in a single PCR well. For additional discussion of the differences between surface-based and solution-based hybridization capture, see Section 3.7.1.

I recommend hybridization capture for urine cfDNA applications where maximum sensitivity is required. Its improvement over alternate methods will be most apparent when paired with an ultrashort PCR target (like the 25 bp target described here). Hybridization capture may be particularly beneficial for highly fragmented cfDNA, such as bacterial, viral, or mitochondrial cfDNA. It may also offer the advantage of increased specificity by removing non-target background DNA, although we have not directly tested this here. A key limitation of hybridization capture is that, unlike silica-based methods, it will only isolate specific targeted sequences. While it is ideal for extraction of a specific diagnostic target, and can be multiplexed to extract multiple targets, it is not suitable for sequencing or other applications requiring broader pulldown of all cfDNA regardless of sequence. Development of capture probes for new targets is straightforward, in this case simply using a truncated version of one of the PCR primers (see Section 3.8).

4.7. Conclusions and recommendations for choice of urine cell-free DNA extraction method

Table 4.3 summarizes the analytical performance of the urine cfDNA extraction methods. My results reveal that extraction methods vary widely in ability to capture the short, dilute cfDNA present in urine. Employing sub-optimal methods may profoundly compromise clinical results due to low recoveries, dependence on urine composition, or PCR inhibition. Overall, hybridization capture and Q Sepharose performed best, with high recovery of short fragments (down to 25 bp and 40 bp, respectively), sensitive detection of dilute fragments, tolerance to varied urine conditions, and resistance to PCR inhibition. As such, these are the two methods expected to perform well in clinical samples and thus recommend for extraction of urine cfDNA. I recommend the surface-based hybridization capture format over the solution-based hybridization capture format due to increased analysis volume and reduced cost. The results of this work will help inform selection of optimal urine cfDNA extraction methods, which, paired with short PCR amplicons, should lead to improved clinical sensitivity and reproducibility of urine cfDNA diagnostics. In addition, the strong analytical performance of hybridization capture compared to alternate urine cfDNA extraction is promising for its performance in clinical urine samples

for TB diagnosis and allowed us to test our hypothesis that improved sample preparation of short, dilute urine cfDNA will lead to an increase in clinical sensitivity. One limitation of this study is that recovery was measured for ssDNA targets rather than dsDNA targets, and it is possible that the methods' performances may differ for dsDNA.

Table 4.3: Summary of analytical performance of urine cfDNA extraction methods.

Urine cfDNA extraction method	Percent recovery	Minimum target length efficiently recovered[a]	Ability to recover low concentrations of short target	Tolerance to varied urine conditions	Resistance to PCR inhibition
Surface-based Hybridization Capture (10 mL)	90–110%	25 bp	Good	Good	Good
Solution-based Hybridization Capture (1 mL)	73–84%	25 bp	Good	Good	Good
Wizard/GuSCN	1.6–17%	40 bp	Poor	Poor	Good
Q Sepharose	8.6–75%	40 bp	Good	Good	Good
Norgen Urine Cell-Free Circulating DNA Purification Kit	30–72%	25 bp	Moderate	Good	Poor
QIAGEN QIAamp Circulating Nucleic Acid Kit	0.20–18%[b]	150 bp	Poor	Good	Good
Thermo Fisher MagMAX Cell-Free DNA Isolation Kit	0.20–66%[b]	80 bp	Poor	Good	Moderate

[a] Efficient recovery defined as >50% of the maximum recovery observed across all lengths for that method.
[b] 25 bp fragment not tested for QIAamp and MagMAX kits.

4.8. Funding Acknowledgements

Research reported in this chapter was supported by University of Washington Coulter Foundation and Global Center for Integrated Health of Women, Adolescents and Children, the National Institute of Allergy and Infectious Diseases of the NIH under award R21AI125975, and the National Science Foundation Graduate Research Fellowship Program.

5. Detection of tuberculosis cell-free DNA in clinical urine samples

This chapter is adapted from

Oreskovic A, Panpradist N, Marangu D, Ngwane MW, Magcaba ZP, Ngcobo S, Ngcobo Z, Horne DJ, Shapiro AE, Wilson DPK, Drain PK, Lutz BR. Diagnosing pulmonary tuberculosis using sequence-specific

purification of urine cell-free DNA. Journal of Clinical Microbiology 59(8):e00074-21 (2021). https://doi.org/10.1128/jcm.00074-21.

This paper describes a clinical study testing the diagnostic accuracy of our TB cfDNA assay in samples from South Africa. Here, Section 5.3 has been added to include the results of an early pilot study that was not included in the above publication and Section 5.5 has been added to include ROC analysis.

5.1. Abstract

Transrenal urine cell-free DNA (cfDNA) is a promising tuberculosis (TB) biomarker but is challenging to detect because of the short length (<100 bp) and low concentration of TB-specific fragments. We aimed to improve the diagnostic sensitivity of TB urine cfDNA by increasing recovery of short fragments during sample preparation. We developed a highly sensitive sequence-specific purification method that uses hybridization capture probes immobilized on magnetic beads to capture short TB cfDNA (50 bp) with 91.8% average efficiency. Combined with short-target PCR, the assay limit of detection was ≤5 copies of cfDNA in 10 mL urine. In a clinical cohort study in South Africa, our urine cfDNA assay had 83.7% sensitivity (95% CI: 71.0–91.5%) and 100% specificity (95% CI: 86.2–100%) for diagnosis of active pulmonary TB when using sputum Xpert MTB/RIF as the reference standard. The detected cfDNA concentration was 0.14–2804 copies/mL (median 14.6 copies/mL) and was inversely correlated with CD4 count and days to culture positivity. Sensitivity was non-significantly higher in HIV-positive (88.2%) compared to HIV-negative patients (73.3%) and was not dependent on CD4 count. Sensitivity remained high in sputum smear-negative (76.0%) and urine LAM-negative (76.5%) patients. With improved sample preparation, urine cfDNA is a viable biomarker for TB diagnosis. Our assay has the highest reported accuracy of any TB urine cfDNA test to date and has the potential to enable rapid non-sputum-based TB diagnosis across key underserved patient populations.

5.2. Introduction

Tuberculosis (TB) is the leading cause of global mortality due to infectious disease, with an estimated 10 million cases and 1.4 million deaths in 2019 (1). An estimated 30% of TB cases remain undiagnosed or unreported, in part due to limitations in rapid diagnostics (1). Current TB tests rely on sputum samples, which are difficult to collect from people living with HIV, severely ill patients, and children, and may not detect extrapulmonary TB (EPTB). Rapid sputum-based tests (e.g., smear microscopy, Xpert MTB/RIF) also have lower sensitivity in these same underserved patient populations, who more often have

paucibacillary TB (7–9). A WHO consensus meeting identified a rapid, non-sputum-based test as one of the highest priority target products for TB diagnostics (2).

Urine is an attractive alternate sample for TB diagnosis because it is easy to collect and poses minimal TB transmission risk. In patients with active TB disease, TB-specific cell-free DNA (cfDNA) fragments are released into the blood, a fraction of which are filtered through the kidneys and excreted in the urine as transrenal cfDNA (3, 4, 52). TB-specific cfDNA has been detected in urine from both HIV-negative and HIV-positive patients with pulmonary TB, but diagnostic sensitivities have been inconsistent (0–79%) (4–6, 17, 59, 60). High variability in methodology and subsequent performance across studies have limited the understanding of TB urine cfDNA and hindered its clinical implementation (42, 82).

Urine cfDNA is a challenging target due to the short length and low concentration of TB-specific fragments. While plasma cfDNA has a peak fragment length of 167 bp, urine cfDNA is more fragmented (3, 45, 56). Recently, new sequencing library preparation methods revealed that very short, formerly-undetectable fragments compose a larger fraction of cfDNA than previously realized (49, 51, 55, 58). The majority of urine cfDNA fragments, regardless of origin, are <100 bp, with peak fragment length ranging from 30–110 bp (45, 54–56). Although the fragment length of TB cfDNA specifically has not yet been characterized, bacterial cfDNA is expected to be especially fragmented (peak <60 bp) because it is less protected by DNA-associated proteins than human cfDNA (53, 55).

The low to moderate sensitivities reported in previous TB urine cfDNA studies are likely due in part to sample preparation and/or amplification methods that are sub-optimal for short urine cfDNA (42, 52). We hypothesized that we could increase the sensitivity of cfDNA-based TB diagnosis by improving recovery of short cfDNA during the DNA extraction step. We developed a highly sensitive sequence-specific purification method that uses DNA probes immobilized on magnetic beads to extract TB-specific cfDNA via hybridization. By combining sequence-specific purification with short-target PCR, we can reliably detect ≤5 copies of short (50 bp) cfDNA in 10 mL urine (see Chapter 3) (162).

Here, we determined for the first time the diagnostic accuracy of our TB cfDNA assay in clinical urine specimens from adults with active pulmonary TB. Our results demonstrate the advantages of our sequence-specific purification approach, contribute to the growing evidence needed to establish urine cfDNA as a TB biomarker, and will serve as the foundation for future clinical studies in expanded populations (e.g., children, individuals with EPTB).

5.3. Pilot clinical study (TB Control Program, Public Health – Seattle & King Co.)

During the early stages of assay development, we conducted pilot testing on a small group of clinical samples collected locally. Initially, I was unable to detect any TB cfDNA in clinical urine samples, but these samples were critical for driving continued assay development, assessing assay improvements and their effect on clinical performance, and generating preliminary clinical data prior to initiating an expanded clinical study. In this section, I summarize the results of this pilot clinical testing and its impact on the development of the final version of the TB urine cfDNA assay used in subsequent clinical testing.

5.3.1. *Methods*

Fifteen HIV-negative adults with suspected pulmonary TB were enrolled at the TB Control Program, Public Health – Seattle & King County (in collaboration with Dr. David Horne). Twelve patients had TB diagnosis confirmed by positive sputum culture. The remaining three patients had negative sputum culture. All patients had ≤7 days of anti-TB treatment. No patients had urine culture positive for TB. After collection, urine was mixed with EDTA to a final concentration of 100 mM and frozen at -80°C for up to three years before analysis. TB cfDNA was extracted from 10 mL urine (n=2–5 replicates per sample, depending on available volume) using an early version of the surface-based hybridization method (Appendix A, Section 9.2.3) and amplified using an early version of the IS6110 PCR without LNA-substituted primers (Appendix A, Section 9.3.1).

5.3.2. *Results of early cfDNA analysis in clinical samples*

I used this first group of samples to gauge the ability of the sequence-specific purification method to detect TB cfDNA in clinical urine samples at various points in the assay development process. All 12 urine samples from sputum culture positive patients were initially negative using the 1 mL solution-based hybridization assay format (Section 3.7.1). Only after scaling up the input sample volume to 10 mL using the surface-based hybridization format did I successfully detect TB-specific cfDNA in clinical urine samples. Using an early version of the 10 mL surface-based hybridization method, I detected TB-specific cfDNA in 75% (n=9/12) of urine samples from patients with TB diagnosis confirmed by sputum culture (Figure 5.1) with a median concentration of 0.27 copies/mL (interquartile range 0.07–4.5 copies/mL, range 0.002–80.0 copies/mL). 58% (n=7/12) were classified as “confident positive,” indicating detection of TB cfDNA across multiple replicates at a concentration roughly equating to at least a single copy of cfDNA in 10 mL urine. 17% (n=2/12) were classified as “weak positive,” one with detection in only a single replicate and one with detection at an unrealistically low concentration (<<1 copy in 10 mL). Of

the three urine samples negative for TB cfDNA, two had a high T_m product (~78°C) that did not match that of the expected native TB target amplicon (75°C). Testing of urine from a healthy control exhibited this same high T_m product. Further troubleshooting revealed nonspecific amplification of residual human genomic DNA, independent of the capture probe (i.e., likely due to nonspecific adsorption of DNA to the bead surface, not nonspecific pull-down by the capture probe). Anecdotally, this problem was more common in samples with higher amounts of cellular debris, which would be expected to have higher human genomic DNA concentrations. I have since resolved this issue by redesigning the PCR primers with LNA substitutions and increasing the PCR annealing temperature to improve specificity (see Section 3.7.4). No TB cfDNA was detected in the urine of two of the three patients with negative sputum culture, or in the urine of a healthy control. One patient with negative sputum culture had detectable TB urine cfDNA, which is likely explained by history of previous TB infection and the possibility of residual cfDNA.

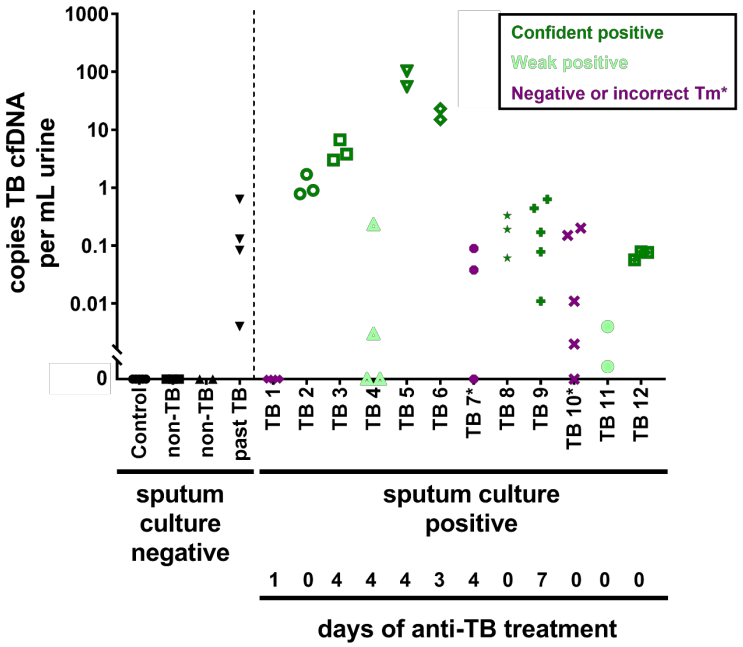


Figure 5.1: Detection of urine cfDNA in 75% of pilot clinical samples from TB Control Program, Public Health – Seattle & King Co. *Urine was collected from HIV-negative patients with suspected TB and ≤7 days anti-TB treatment and stored up to 3 years at -80°C with 100 mM EDTA. TB urine cfDNA was extracted from 10 mL urine using an early version of the surface-based hybridization method, which has since been improved and can now detect 4 – 8X more cfDNA. Samples were classified as confident positive (detection in all replicates), weak positive (detection in single replicate or at unrealistically low concentration), or weak positive (no TB cfDNA detected or incorrect T_m). Each data point represents an individual extraction replicate (n=2-5 replicates per sample, depending on available volume). *Indicates T_m higher than that of expected amplicon, which was due to nonspecific amplification of residual human genomic DNA and has now been resolved. These samples were classified as negative.*

5.3.3. *Limitations of early cfDNA analysis in clinical samples*

While the results of this preliminary clinical testing were promising, and suggested that TB-specific cfDNA could be detected in urine with sensitivity greater than or similar to previous studies, there were several limitations. The concentrations of detected cfDNA were low, in some cases below a single copy in 10 mL urine. It is not clear whether the low concentrations detected were due to low concentrations of target present in the urine samples and/or deficiencies in the capture method. The clinical concentration range for TB urine cfDNA has not yet been firmly established, but is expected to be low (median 6.8 copies/mL; range 1–40.7 copies/mL for 7 patients reported by Labugger *et al.*) (5). The samples analyzed here were also stored for up to three years prior to analysis, so it is possible that cfDNA may have degraded over time. The optimal storage conditions for urine cfDNA are unknown, and urine cfDNA stability may vary across patient populations (87–92). A few samples had notably higher concentrations of TB cfDNA, but we do not have paired clinical data to determine if these patients had higher bacterial load. In subsequent studies, additional clinical data to correlate with urine cfDNA concentration were included. Because the samples were tested sporadically over time as a benchmark of the hybridization capture method's sensitivity to detect clinical TB urine cfDNA, I also used an inconsistent protocol as I made incremental changes across the duration of sample testing.

5.3.4. *Further improvements made to assay after pilot testing*

After the completion of this pilot study, I made further improvements to the sequence-specific purification method, leading to a 4- to 8-fold increase in detected cfDNA. The final optimized protocol, employed in later clinical testing, has >90% average recovery (see Figure 5.3), but the recovery of the early method used here (Appendix A, Section 9.2.3) was more variable, with recovery as low as 23%. Key improvements included switching from single to dual biotin modified probes (Section 3.7.2), improving the retention of beads during wash steps by concentrating to a 1 mL volume prior to washing (Appendix B, Section 10.1.6), improving denaturation (Appendix B, Section 10.1.1), and adding a second capture probe to target both strands of dsDNA (Section 3.7.3). I also modified the primer design and PCR cycling to eliminate the nonspecific amplification of residual cell-associated human genomic DNA observed in this pilot study (Section 3.7.4). Together, these improvements should improve clinical performance by enabling more confident detection of low concentration samples. Specificity testing was also limited during pilot clinical testing but was expanded in follow-up clinical work. Ultimately, this preliminary clinical work, while limited in scope and by the methods used, motivated continued assay development and gave confidence to pursue an expanded clinical study.

5.4. Clinical study: Adult pulmonary TB (Edendale Hospital, South Africa)

In this section, I describe an expanded clinical cohort study to determine the accuracy of our TB urine cfDNA for diagnosis of adult pulmonary TB.

5.4.1. *Methods*

Study design and participants. The design of this study is summarized in Figure 5.2. Participants were consecutively enrolled at Edendale Hospital in Pietermaritzburg, South Africa (approved by the University of KwaZulu-Natal Biomedical Research Ethics Committee, #BE475/18). Adults (≥18 years old) meeting the case definition for active pulmonary TB were recruited. As a control group, adult inpatients at Edendale Hospital with excluded diagnosis for active TB disease were concurrently enrolled. The characteristics of the South Africa control group closely matched the TB-positive study population (i.e., patients seeking hospital care, many HIV-positive, likely included participants with latent TB and Bacillus Calmette–Guérin vaccination). Because the TB burden at the enrollment site in South Africa is high, however, there was a risk of enrolling patients with undiagnosed TB in the control group. To differentiate any potential false positives due to undiagnosed TB, a separate healthy control group was enrolled at the University of Washington, Seattle, USA (approved by the University of Washington Institutional Review Board, #48840). Participants were adults (≥18 years old) with low risk of TB exposure, as defined by birth in a country with low TB risk and no history of diagnosis of latent TB, treatment for active or latent TB, or living with an individual with active TB. All participants provided written informed consent. Samples were de-identified prior to testing at the University of Washington where this study was conducted.

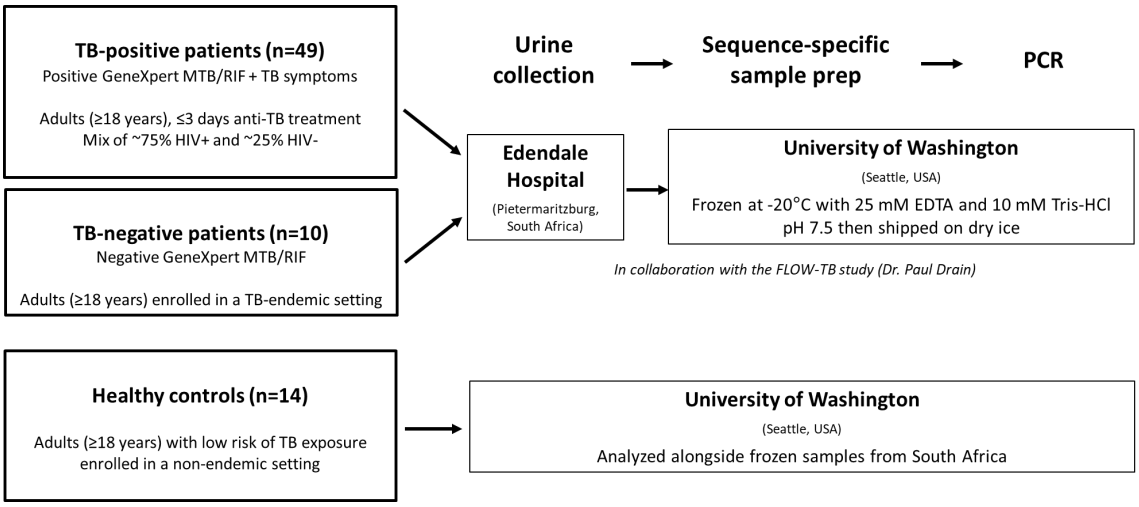


Figure 5.2: Study design for adult pulmonary TB.

Case definitions. TB-positive participants (South Africa) were defined as those with a positive sputum Xpert MTB/RIF (Cepheid, Sunnyvale, CA, USA) result within 5 days of enrollment and the presence of one or more TB symptom (fever, night sweats, cough, and weight loss). Xpert MTB/RIF rather than Xpert MTB/RIF Ultra was used as the reference standard because Xpert MTB/RIF Ultra cartridges were not available at Edendale Hospital, a regional healthcare facility, at the time of participant enrollment (February – July 2019). Patients with >72 hours of anti-TB treatment were excluded. TB-negative participants (South Africa) had a primary diagnosis other than TB and no clinical suspicion for TB. Healthy controls (USA) were recruited on a volunteer basis and were not seeking medical care.

Clinical data, sputum testing, and urine lipoarabinomannan (LAM) testing. Clinical data were collected for all participants enrolled in South Africa. These included date of birth, gender, weight, height, self-reported TB history, presence of TB symptoms (fever, night sweats, cough, weight loss), TB treatment duration, HIV test result, and $CD4^+$ cell count (for participants living with HIV only). Expectorated sputum was submitted to the South African National Health Laboratory System (NHLS) for confirmatory solid and liquid mycobacterial culture and AFB smear microscopy. Mycobacterial culture was performed at the NHLS Provincial TB Reference Laboratory using both Middlebrook 7H11 solid agar medium and the liquid BACTEC mycobacterial growth indicator tube (MGIT) 960 system (BD, Franklin Lakes, NJ, USA) for each sputum sample. Cultures were incubated for up to 42 days. Culture plates were read at 3 and 6 weeks, and *M. tuberculosis* was identified from solid or liquid cultures using niacin and nitrate testing. Participants were considered culture-positive with growth from either the solid or liquid culture. Smear microscopy was performed in the NHLS Laboratory at Edendale Hospital on decontaminated samples using both Ziehl-Neelsen and Auramine stains and considered positive if either stain revealed AFB. Urine (60 µL) was tested using Alere Determine TB LAM Ag (Abbott Laboratories, Chicago, USA). Collection of clinical data, sputum testing, and urine LAM testing were not done for healthy controls enrolled in the USA.

Urine collection and storage for cfDNA analysis. Participants were asked to provide a convenience urine sample (50–200 mL) at the time of enrollment and/or an early morning first-void sample the morning after enrollment. The samples were not obtained mid-stream. As soon as possible after collection, urine was mixed with EDTA and Tris-HCl pH 7.5 to final concentrations of 25 mM and 10 mM, respectively. Urine was stored in DNA LoBind tubes (Eppendorf, Hamburg, Germany) at -20°C at the collection site until shipping. Samples were shipped on dry ice to the University of Washington, where they were stored at -80°C until analysis. Immediately before analysis, urine was thawed at 37°C and centrifuged at

8,000g for 5 minutes to pellet cell debris. The cell-free urine supernatant was transferred to new 15 mL DNA LoBind tubes (Eppendorf) and characterized using Fisherbrand 10-SG Urine Reagent Strips (Thermo Fisher Scientific, Waltham, MA, USA) to measure the levels of glucose, bilirubin, ketone, specific gravity, blood, pH, protein, urobilinogen, nitrite, and leukocytes.

Purification of TB cfDNA from urine using sequence-specific hybridization capture. cfDNA was extracted in triplicate from 10 mL urine samples (3x10 mL samples for each participant), with individual replicates for each participant processed on separate days. TB-specific urine cfDNA was extracted using our in-house sequence-specific hybridization capture method, as described previously (Chapter 3) (162). We have published a detailed, user-ready protocol at http://dx.doi.org/10.17504/protocols.io.bep4jdqw.

Immobilization of capture probes on magnetic beads: Dynabeads MyOne Streptavidin C1 (Thermo Fisher) were washed three times with an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8.0, 0.05% (v/v) Tween-20) and resuspended in an equal volume of high salt wash buffer. Dual biotinylated capture probes BP1 and BP2 (Table 5.1), targeting opposite strands of the double-stranded IS*6110* target region, were pre-mixed in low-EDTA TE buffer to a concentration of 50 µM each. 25 pmol of each probe (0.5 µL of probe mix) per 50 µL bead equivalent was added to the beads. The beads were immediately vortexed and rotated for 15 minutes at room temperature to immobilize capture probes on the beads. The beads were washed three times with an equal volume of high salt wash buffer and resuspended in an equal volume of high salt wash buffer.

Table 5.1 Probe, primer, and target sequences.[a]

Oligo	Sequence
Dual biotinylated capture probe #1 (BP1)	5'-/52-Bio/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/-3'
Dual biotinylated capture probe #2 (BP2)[b]	5'-/52-Bio/AAAAAAAAAAAAAAAAAAAACCCTGCCCAGGTCGA/3SpC3/-3'
Forward primer	5'- CGAACCCTGCCCAGGTCGA-3'
Reverse primer	5'- GTA+GCAGA+CCTCACCTATGTGT-3'
IS6110 target (40 bp)	5'- CGAACCCTGCCCAGGTCGACACATAGGTGAGGTCTGCTAC-3'
IS6110 reverse complement (40 bp)	5'-GTAGCAGACCTCACCTATGTGTCGACCTGGGCAGGGTTCG-3'
Synthetic positive control (50 bp)	5'- CGAACCCTGCCCAGGTCGACA**CCATTCAACA**CATAGGTGAGGTCTGCTAC-3'

Synthetic positive control reverse complement (50 bp)	5’-GTAGCAGACCTCACCTATG**TGTTGAATGG**TGTCGACCTGGGCAGGGTTCG-3’

[a] /52-Bio/ indicates dual biotin modification; /3SpC3/ indicates carbon spacer; “+G” and “+C” indicate locked nucleic acid (LNA) bases. Target-specific probe binding sequences are underlined. A synthetic spacer region introduced to differentiate the synthetic positive control from the endogenous Mycobacterium tuberculosis complex-specific target sequence (IS6110) is bolded. All DNA sequences were ordered HPLC-purified from Integrated DNA Technologies (Coralville, IA, USA).

[b] BP2 targets the opposite strand (reverse complement) of the DNA strand targeted by BP1.

Sequence-specific capture of TB cfDNA: 2.5 mL of 5 M NaCl (final concentration 1 M), 127 µL of 10% (v/v) Tween-20 (final concentration 0.1%), and 50 µL of prepared beads were added to each 10 mL urine sample and gently mixed. The samples were denatured in a heat block set to 120°C for 15 minutes, so that the urine reached a temperature of >90°C. TB cfDNA was hybridized to capture probes by rotation at room temperature for 30 minutes.

Washing to remove urine and non-target DNA: Samples were centrifuged for 5 minutes at 5000g to pellet beads. All but 1 mL of supernatant was removed and discarded. Beads were resuspended in the remaining supernatant and transferred to 1.5 mL DNA LoBind tubes (Eppendorf). The tubes were placed on an Invitrogen Dynamag-2 magnetic rack (Thermo Fisher) for 1 minute and the remaining supernatant was discarded. The beads were washed twice with 1 mL high salt wash buffer and once with 1 mL low salt wash buffer (15 mM NaCl, 10 mM Tris-HCl pH 8.0). For each wash, the tubes were inverted 10–20 times, or until no bead aggregate was left on the tube walls, briefly spun down, and placed on the magnetic rack for 1 minute before removing the wash buffer. After the final wash step, beads were spun down and any residual buffer was removed.

Elution of purified TB cfDNA: Purified TB cfDNA was eluted using 20 µL of freshly prepared 20 mM NaOH. Beads were mixed with NaOH by vortexing, briefly spun down, and placed on the magnetic rack. The eluate was transferred directly to PCR wells and partially neutralized with 3.5 µL 100 mM HCl.

Amplification of TB-specific urine cfDNA using short-target PCR. The entire output (~24 µL) from each 10 mL urine sample was analyzed in a single PCR well. Each 50 µL reaction contained 1.25 U OneTaq Hot Start DNA Polymerase (New England Biolabs [NEB, Ipswitch, MA, USA]), 1X OneTaq GC Reaction Buffer (NEB; 80 mM Tris-SO_4, 20 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 5% glycerol, 5% dimethyl sulfoxide, 0.06% IGEPAL CA-630, and 0.05% Tween-20, pH 9.2), 0.8 mM dNTPs (NEB), 0.4X EvaGreen (Biotium, Fremont, CA, USA), 200 nM forward primer (Table 5.1), and 200 mM reverse primer (Table 5.1). qPCR was carried out in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) using an initial incubation period of 94°C for 3 minutes, followed by 45 amplification cycles of 94°C for 30

seconds, 64°C for 30 seconds, and 68°C for 1 minute. Quantification cycle (C_q) values were determined using the CFX Maestro software version 1.1 (Bio-Rad Laboratories) at a threshold of 500 RFU and recovered copies were calculated using a standard curve run with each PCR. PCR products were confirmed by post-amplification melt curve analysis from 65°C to 95°C in 0.5°C increments every 5 seconds.

Criteria for positive samples. Criteria for positive samples were set prior to analysis, and the assay operator was blinded to the clinical status of the samples until after processing was complete and sample calls were made. Individual replicates were ruled as positive if ≥1 single-stranded DNA copy of TB cfDNA was detected and the melt temperature (T_m) matched that of the expected native TB amplicon (76°C). Individual replicates were ruled as negative if insufficient TB cfDNA was detected (<1 copy of single-stranded DNA) and/or the T_m did not match that of the expected native TB amplicon. To allow for expected occasional PCR drop-out of low-concentration samples, a sample was ruled as positive if at least two of three replicates, analyzed on different days, were positive based on the above definitions. Samples with zero or one positive replicates were ruled as negative.

PCR primer design, controls, and precautions to prevent false positives. We designed PCR primers targeting the insertion sequence IS6110, which is an established target for TB diagnosis present at multiple, variable copy number across ~99% of *Mycobacterium tuberculosis* complex strains (129). Our PCR primers (Table 5.1) amplify a short 40 bp target within a subregion of IS*6110* that is conserved and specific to the *Mycobacterium tuberculosis* (MTB) complex (130). To avoid false positives due to nonspecific amplification, we used locked nucleic acid (LNA) bases to precisely match the T_m of the primers and selected an annealing temperature (T_a) slightly above the primer T_m to encourage specific amplification without compromising amplification efficiency. The final primer set and optimized PCR conditions result in an average PCR efficiency of 97.1% (95% CI: 95.7–98.6%) and no amplification of no template controls (NTCs) or 10 ng of human genomic DNA up to at least 45 cycles. To avoid false positives due to contamination, we maintained good laboratory practices to limit contamination (e.g., separating pre- and post-PCR rooms, regular decontamination of work surfaces and pipettes, sterile filtered pipette tips, aliquoting reagents into single-use volumes). In addition, we designed the synthetic positive control (50 bp, used as a spike-in control during extraction and for PCR standard curves) to be amplifiable by the same primer pair but distinguishable from the endogenous TB target sequence (40 bp) by melt analysis (Table 5.1) so that any potential contamination with the positive control would not result in false positives. Every experiment included a positive control (pooled TB-negative urine spiked

with 10^3 copies of double-stranded DNA synthetic positive control template [Table 5.1]) and negative control (water without spiked target) that were run throughout the entire extraction process alongside clinical urine samples. PCR NTCs (n=3) were also included in each experiment.

Statistical analysis. Sensitivity and specificity were calculated using a positive sputum Xpert MTB/RIF result and the presence of one or more TB symptoms as the reference standard. 95% confidence intervals for sensitivity and specificity were calculated using the hybrid Wilson/Brown method. Sensitivities were compared across groups using Fisher's exact test. The detected cfDNA concentration was calculated based on the sample means of cfDNA-positive samples. Detected cfDNA concentrations were compared across groups using the Mann-Whitney test. Correlations with detected cfDNA concentration were assessed using Spearman's correlation coefficient. Odds ratios were compared to a value of 1 using Fisher's exact test, with 95% confidence intervals determined using the Baptista-Pike method. All statistical analysis was conducted using GraphPad Prism v8.1.2 (San Diego, CA, USA) with a significance level of 0.05.

5.4.2. *Results*

Analytical performance of sequence-specific TB cfDNA assay. An overview of our cfDNA assay is shown in Figure 5.3A-B. Recovery of the spiked-in positive control was 91.8% (95% CI: 75.9–107.6%) across 15 independent experiments (Figure 5.3C).

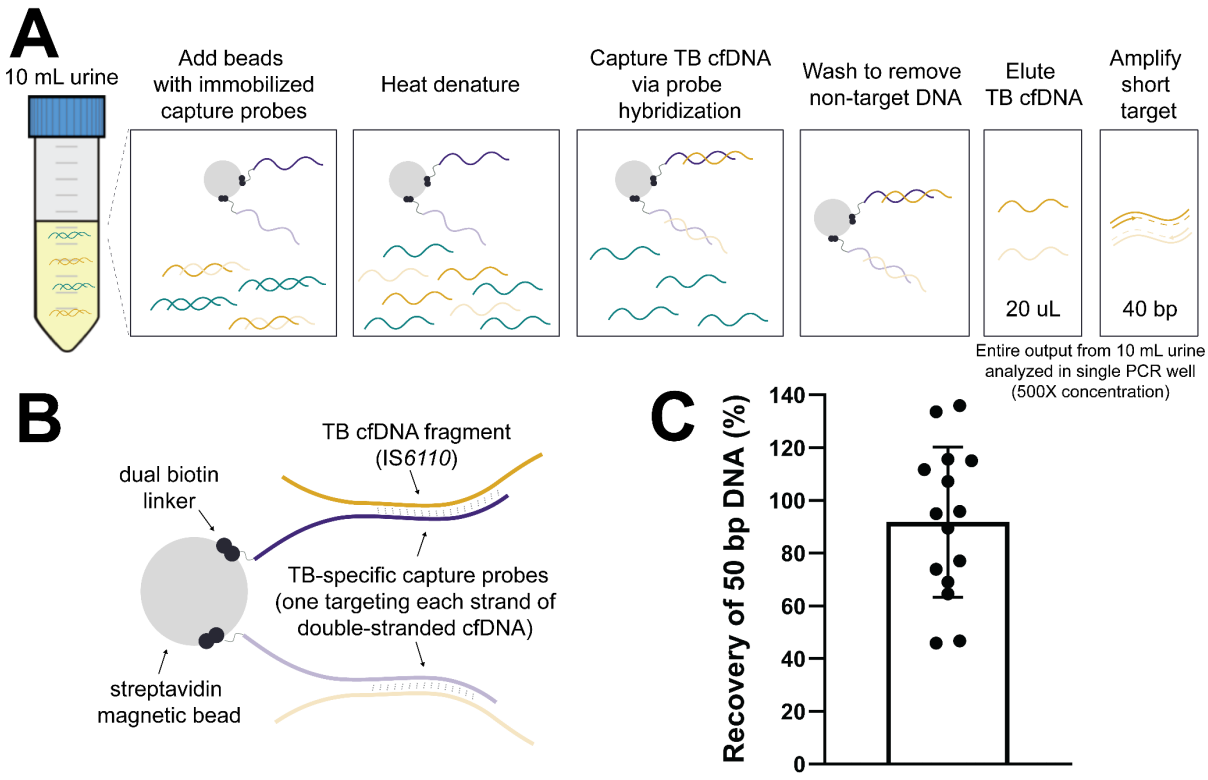


Figure 5.3: Capture and detection of short cfDNA fragments in urine using sequence-specific purification and short-target PCR. (A) *Overview of sequence-specific purification and short-target PCR (40 bp) protocol for TB cfDNA.* ***(B)*** *Schematic illustrating details of capture probe design targeting the MTB complex-specific insertion element IS6110.* ***(C)*** *Near complete recovery of short TB-specific DNA spiked into urine using sequence-specific purification. A positive control (10^3 copies of 50 bp double-stranded DNA in 10 mL pooled urine) was included throughout each experiment alongside clinical samples. The recovery of the positive control was calculated as a percentage of the input (mean ± SD, n=15). Key design features, assay optimization, and additional analytical characterization of our sequence-specific purification method for cfDNA are reported in (Chapter 3)* (162).

Study participants. We enrolled 73 participants across two sites: Edendale Hospital in South Africa (49 TB-positive, 10 TB-negative) and the University of Washington in the USA (14 healthy controls with low risk of TB exposure). Participant demographic and clinical data are summarized in Table 5.2.

Table 5.2: Summary of study participants.

		TB-positive[a] patients (South Africa)	TB-negative controls (South Africa)	Healthy controls (USA)
Total [n]		49	10	14
Gender [n (%)]	Female	22 (44.9%)	8 (80.0%)	5 (35.7%)
	Male	27 (55.1%)	2 (20.0%)	9 (64.3%)
Age, years [median (IQR)]		38 (31 – 48)	32 (30.25 – 37.25)	26.5 (24 – 28.75)
Height, m [median (IQR)]		1.68 (1.65 – 1.70)	1.60 (1.59 – 1.68)	n/a
Weight, kg [median (IQR)]		55 (48 – 63)	57 (52 – 70)	n/a
HIV status [n (%)]	HIV-positive	34 (69.4%)	10 (100.0%)	n/a
	HIV-negative	15 (30.6%)	0 (0.0%)	n/a
$CD4^+$ T-cell count ($cells/mm^3$)[b] [median (IQR)]		85 (42 – 345)	318 (257 – 746)	n/a
History of prior TB [n (%)]	No	32 (65.3%)	6 (60.0%)	14 (100.0%)
	Yes	17 (34.7%)	4 (40.0%)	0 (0.0%)
TB treatment status [n (%)]	Treatment-naïve	36 (73.5%)	n/a	n/a
	Some treatment[c]	13 (26.5%)	n/a	n/a
TB culture result [n (%)]	Culture-positive	34 (69.4%)	0	n/a
	Culture-negative	8 (16.3%)	9 (90.0%)	n/a
	No culture result	7 (14.3%)	1 (10.0%)	n/a
Days to culture positivity[d] [median (IQR)]		13.5 (8.0 – 17.0)	n/a	n/a
AFB smear result [n (%)]	Smear-positive	19 (38.8%)	0 (0.0%)	n/a
	Smear-negative	25 (51.0%)	5 (50.0%)	n/a
	No smear result	5 (10.2%)	5 (50.0%)	n/a
Alere urine LAM result [n (%)]	LAM-positive	15 (30.6%)	0 (0.0%)	n/a
	LAM-negative	34 (69.4%)	10 (100.0%)	n/a

[a] TB-positive patients were defined as those with a positive Xpert MTB/RIF result and the presence of one or more TB symptoms.
[b] CD4 count was measured for HIV-positive patients only.
[c] All participants had ≤72 hours of treatment.
[d] For mycobacterial growth indicator tube (MGIT) culture only.

Sensitivity and specificity of sequence-specific TB cfDNA assay. A summary of the diagnostic accuracy of our TB urine cfDNA assay is given in Table 5.3. Full results for each participant, including paired clinical data, urine characteristics, and detected cfDNA concentration, are given in Appendix D, Table 12.2 and

Table 12.3. Using sputum Xpert MTB/RIF as the reference standard, we detected TB-specific urine cfDNA with 83.7% sensitivity (n=41/49; 95% CI: 71.0 – 91.5%) and 100% specificity (n=24/24; 95% CI: 86.2–100%). No TB-specific cfDNA was detected in the urine of TB-negative controls in South Africa (n=10) or healthy controls in the USA (n=14).

Table 5.3: Sensitivity and specificity of TB urine cfDNA assay.

		cfDNA-positive/ TB-positive (n/N)	Sensitivity (95% CI[a]) (%)	P-value (comparing sensitivity across subgroups)[b]	cfDNA-negative/ TB-negative (n/N)	Specificity (95% CI[a]) (%)
Total[c]		41/49	83.7 (71.0 – 91.5)		24/24	100 (86.2 – 100)
HIV status	HIV-positive	30/34	88.2 (73.4 – 95.3)	0.23	10/10	100 (72.3 – 100)
	HIV-negative	11/15	73.3 (48.1 – 89.1)		0/0	n/a
CD4$^+$ count[d]	≤200 cells/mm^3	20/22	90.9 (72.2 – 99.4)	0.60	1/1	100 (51.0 – 100)
	>200 cells/mm^3	10/12	83.3 (55.2 – 97.0)		9/9	100 (70.1 – 100)
Sputum culture result	Positive	30/34	88.2 (73.4 – 95.3)	0.32	0/0	n/a
	Negative	6/8	75.0 (40.9 – 95.6)		9/9	100 (70.1 – 100)
AFB sputum smear result	Positive	19/19	100 (83.2 – 100)	0.029*	0/0	n/a
	Negative	19/25	76.0 (56.6 – 88.5)		5/5	100 (56.6 – 100)
Alere urine LAM result	Positive	15/15	100 (79.6 – 100)	0.087	0/0	n/a
	Negative	26/34	76.5 (60.0 – 87.6)		10/10	100 (72.3 – 100)
TB treatment status	Treatment-naïve	28/36	77.8 (61.9 – 88.3)	0.090	10/10	100 (72.3 – 100)
	Some treatment[e]	13/13	100 (77.2 – 100)		0/0	n/a
Gender[b]	Female	18/22	81.8 (61.5 – 92.7)	>0.99	13/13	100 (77.2 – 100)
	Male	23/27	85.2 (67.5 – 94.1)		11/11	100 (74.1 – 100)

** Indicates P < 0.05*

[a] 95% confidence intervals for sensitivity and specificity were calculated using the hybrid Wilson/Brown method.

[b] P-values comparing sensitivity across groups were calculated using Fisher's exact test for relative risk ratios.

[c] Healthy controls enrolled in the USA were included in total specificity and gender-specific specificity, but not in the remaining specificities.

[d] CD4 count was measured for HIV-positive patients only.

[e] All participants had ≤72 hours of treatment.

Sensitivity was non-significantly higher in HIV-positive patients compared to HIV-negative patients (88.2% [n=30/34; 95% CI: 73.4–95.3%] vs. 73.3% [n=11/15; 95% CI: 48.1–89.1%]; p=0.23). Sensitivity was similar in HIV-positive patients with CD4 counts of ≤200 and >200 cells/mm^3 (90.9% [n=20/22; 95% CI: 72.2–99.4%] vs. 83.3% [n=10/12; 95% CI: 55.2–97.0%]; p=0.60). Sensitivity was 77.8% (n=28/36; 95% CI: 61.9–88.3%) in treatment-naïve patients. If a positive sputum culture result was also required for the TB case definition, total sensitivity increased to 88.2% (n=30/34; 95% CI: 73.4–95.3%). We detected TB-specific cfDNA in the urine of all patients with positive AFB sputum smear (n=19/19) and/or Alere urine

LAM (n=15/15) results. Sensitivity remained high in smear-negative (76.0% [n=19/25; 95% CI: 56.6–88.5%]) and LAM-negative (76.5% [n=26/34; 95% CI: 60.0–87.6%]) patients. The reduction in sensitivity for smear-negative patients compared to smear-positive patients was significant (p=0.029), while that for LAM-negative compared to LAM-positive patients was not (p=0.087). See Appendix D, Table 12.6 for statistical analysis of sensitivity across subgroups. Due to the relatively small sample size of this study, the power of these subgroup comparisons was limited and the statistical analyses should be interpreted accordingly. Subsequent, more highly powered studies are needed to conclusively compare cfDNA detection sensitivity across subgroups.

We evaluated factors associated with TB cfDNA positivity and found that only a positive AFB smear result was significantly associated with a positive TB cfDNA result (odds ratio >1, p=0.029). HIV status, CD4 count, treatment status, culture result, and LAM result were not significantly associated with a positive urine cfDNA result (Appendix D, Table 12.4).

Quantification of TB-specific cfDNA concentration in urine. The detected TB-specific cfDNA concentration was variable and skewed towards low concentrations, with a median of 146 copies in 10 mL urine (Table 5.4). The cfDNA concentration had a moderate inverse correlation with CD4 count (-0.43 [95% CI: -0.68 to -0.10; p=0.011]) (Figure 5.4A) and days to culture positivity (-0.36 [95% CI: -0.64 to -0.0060; p=0.041]) (Appendix D, Figure 12.1A and Table 12.5), but no significant correlation with days of anti-TB treatment (1-3 days), AFB smear score, or Alere urine LAM score (Appendix D, Figure 12.1B-D and Table 12.5). The cfDNA concentration was higher in patients with some treatment compared to treatment-naïve patients (p=0.045) and in urine LAM-positive patients compared to urine LAM-negative patients (p=0.0045), but was not significantly different in patients grouped by HIV status, CD4 count, culture result, smear result, or gender (Figure 5.4B).

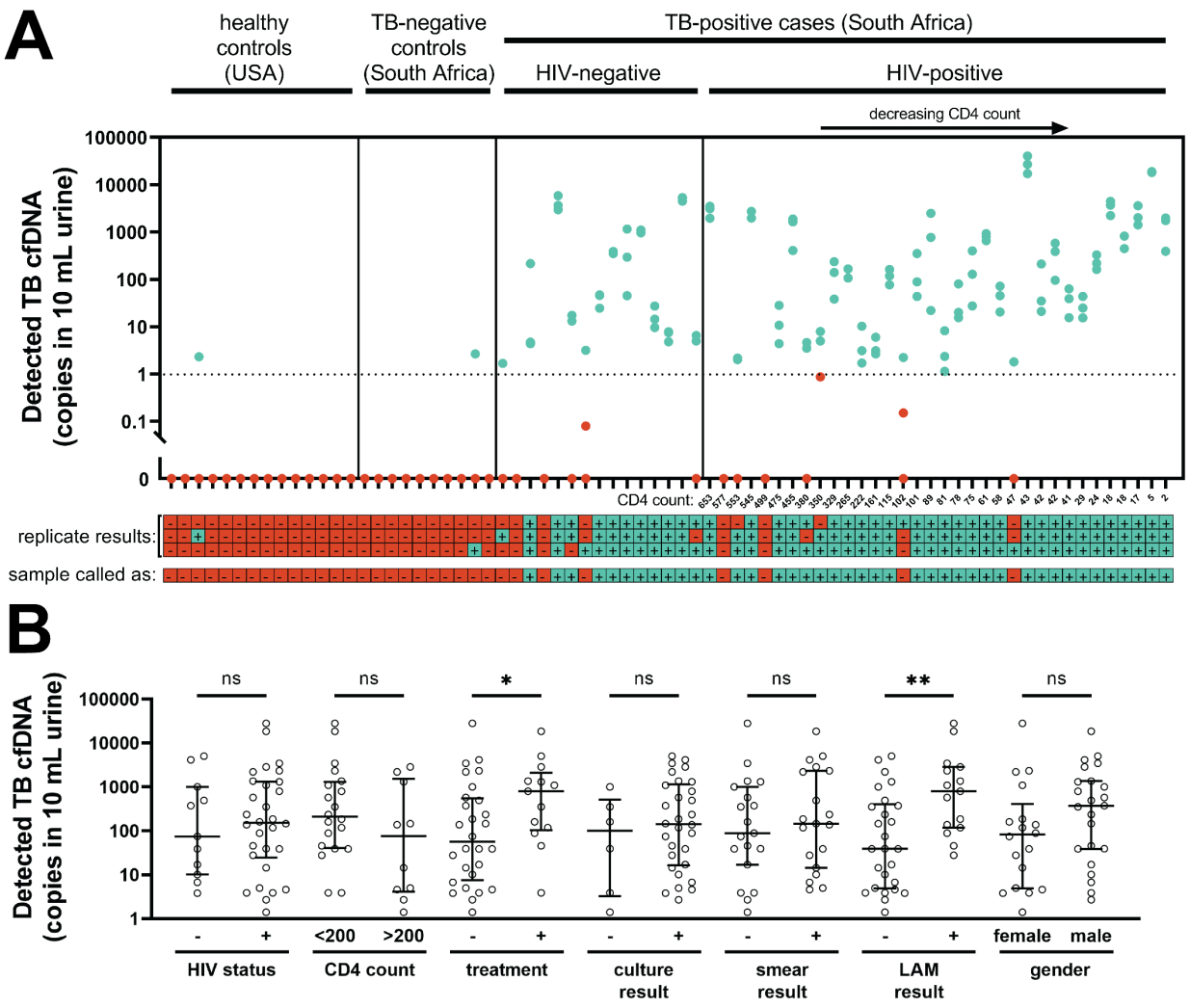


Figure 5.4: Detected concentrations of TB-specific urine cfDNA. (A) *Concentration of TB cfDNA detected in each participant's urine, stratified by HIV status and ranked by CD4 count. There was a moderate inverse correlation between CD4 count and detected TB cfDNA concentration (Spearman's ρ = -0.43 [95% CI: -0.68 to -0.10], p = 0.011), but TB cfDNA could be detected regardless of HIV status and CD4 count. Each dot represents one of three replicates processed on different days for each sample. Note that dots representing replicates with similar detected concentrations may overlap. Replicates called as positive are shown in cyan and replicates called as negative are shown in red. The dashed line indicates the 1 copy per 10 mL threshold used to define positive replicates. The legend below the plot indicates cfDNA detection status by replicate (considered positive if ≥1 copy of single-stranded DNA was detected with melt temperature matching that of the expected IS6110 amplicon) and by sample (considered positive if ≥2 of 3 replicates were positive).* ***(B)*** *Comparison of detected TB cfDNA concentration across groups (bars indicate median and IQR of sample means of cfDNA-positive samples). The detected TB cfDNA concentration was significantly higher in patients with some treatment compared to treatment-naïve patients, and in LAM-positive patients compared to LAM-negative patients (* indicates Mann-Whitney P < 0.05; ** indicates Mann-Whitney P < 0.01), but was not affected by HIV status, CD4 count, culture result, smear result, or gender (ns indicates not significant). See Appendix D, Table 12.6 for calculated P-values for each comparison.*

Table 5.4: Detected concentrations of TB-specific urine cfDNA.

		Median (IQR)[a] [copies in 10 mL urine]	Range[a] [copies in 10 mL urine]
Total		146 (17 – 1092)	1.4 – 28044
TB treatment status[b]	Treatment-naïve	57 (7.6 – 557)	1.4 – 28044
	Some treatment[c]	796 (104 – 2111)	3.9 - 18543
Alere urine LAM result[d]	Positive	796 (119 – 2851)	28 – 28044
	Negative	39 (4.9 – 404)	1.4 - 5021

[a] Concentration for each sample is given as copies of single-stranded DNA in 10 mL urine, which was calculated based on the mean across n=3 technical replicates.
[b] Detected cfDNA concentration was higher in patients with some treatment (Mann-Whitney, p=0.045).
[c] All participants had ≤72 hours of treatment.
[d] Detected cfDNA concentration was higher in patients with a positive urine LAM result (Mann-Whitney, p=0.0045).

5.4.3. *Discussion*

Improved detection of short cfDNA in urine using sequence-specific purification. To maximize sensitivity of TB urine cfDNA detection, it is essential to use methods designed to detect low concentrations of short fragments (42, 52, 82). Decreasing the minimum target length is expected to improve cfDNA diagnostic sensitivity (4, 78, 79, 81). For example, decreasing PCR amplicon length by a modest 10 bp (49 bp to 39 bp) led to more than 10-fold increase in detected TB-specific cfDNA (81). Critically, in addition to amplifying short targets, sample preparation methods must also extract short cfDNA from urine with high efficiency. Conventional silica-based DNA extraction methods have reduced recovery of short fragments and are thus not optimal for urine cfDNA (73).

We aimed to increase the diagnostic sensitivity of TB urine cfDNA detection by improving recovery of short cfDNA during the DNA extraction step. We developed a sequence-specific purification method that uses hybridization capture probes immobilized on magnetic beads to extract short cfDNA with high analytical sensitivity. We have previously demonstrated that sequence-specific purification improves recovery of short cfDNA fragments (25–150 bp) from urine compared to alternate urine cfDNA extraction methods, including a protocol used for TB urine cfDNA (Chapter 4) (127). In a recent paper detailing our sequence-specific purification method, we described its key features, provided a user-ready protocol, and thoroughly characterized its analytical performance (Chapter 3) (162). In this study, our sequence-specific approach recovered nearly all target-specific 50 bp positive control DNA (91.8% average recovery) and was in some clinical specimens able to detect down to a single copy of TB cfDNA in 10 mL urine.

Diagnostic accuracy and comparison to previous TB urine cfDNA studies. In this study, we tested our sequence-specific TB cfDNA assay in clinical urine specimens for the first time. The sensitivity and

specificity of our assay for diagnosis of active pulmonary TB were 84% and 100%, respectively, the highest reported diagnostic accuracy of a TB cfDNA test. For comparison, previous TB urine cfDNA studies are summarized in Table 5.5 (4–6, 17, 59). For a comprehensive review of previous studies, refer to (42).

Table 5.5: Comparison to previous studies targeting urine cfDNA for pulmonary TB diagnosis.[a]

First author	Year	DNA extraction		PCR amplicon length	Effective urine volume analyzed[b]	Proportion of smear-negative participants[c]	Sensitivity		Specificity
		Method	Designed for short urine cfDNA?				HIV-negative patients	HIV-positive patients	
Cannas (4)	2008	Wizard silica resin	Claimed, but we have identified limitations (<35% recovery of ≤150 bp DNA)[d]	129/67 bp (nested)	0.35 mL	2/43	79% (34/43) [95% HIV-negative]		100% (23/23)
Peter (17)[e]	2012	Xpert cartridge	No	192 bp (Xpert)	NR	NR	NA	8% (3/38)	NA
Fortún (60)	2014	NR	No	NR (AMTD)	NR	NR	18% (5/28)		NA
Bordelon (59)	2017	Dynabeads MyOne Silane	No	67 bp	0.1 – 0.5 mL	NR	0% (0/33)		0% (0/31)
Labugger (5)	2017	Unspecified silica resin[f]	Claimed (33% recovery of 75 bp DNA)[g]	38 bp	1.6 mL	4/10	64% (7/11)	NA	100% (8/8)
Patel (6)	2018	Unspecified silica resin[f]	Claimed (~50% recovery of gDNA)[h]	38 bp	NR	45/175	40% (36/90)	45% (38/84)	89% (210/237)
Oreskovic (163)	2021	Sequence-specific purification	Yes (>90% recovery of 50 bp DNA)	40 bp	10 mL	30/49	73% (11/15)	88% (30/34)	100% (24/24)

NR indicates not reported. NA indicates not applicable. Xpert indicates Xpert MTB/RIF assay (Cepheid). AMDT indicates Amplified Mycobacterium Tuberculosis Direct Test (Hologic). gDNA indicates TB genomic DNA.

[a] Included studies specifically targeted urine cell-free DNA for pulmonary TB diagnosis. Studies analyzing cell-associated DNA (18, 19, 72, 63, 64, 66–71)*, plasma cfDNA* (61, 62, 164)*, or EPTB were excluded. Studies targeting plasma cfDNA were excluded here but reported 65%/93%* (62)*, 29%/100%* (62)*, and 45%/67%* (61) *sensitivity/specificity.*

[b] Calculated based on the urine input volume, elution volume, and PCR input volume.

[c] Given as proportion of TB-positive participants who had a negative sputum smear microscopy result. cfDNA detection sensitivity may be lower in smear-negative patients compared to smear-positive patients.

[d] We found that the Wizard silica method had low recovery of short fragments (<35% for 40 – 150 bp) and was dependent on urine composition (i.e., pH and non-target DNA concentration) (127)*.*

[e] Peter et al. tested both whole urine and the insoluble fraction of urine concentrated by centrifugation. Only results from whole urine testing are listed here because cfDNA is expected to be in the soluble fraction of centrifuged urine.

[f] Labugger et al. and Patel et al. used a silica resin-based method similar to that used by Cannas et al., but did not specify the resin or binding buffer.

[g] *Estimated based on reported average of 1.6 cycle delay for extraction calibration curves compared to PCR calibration curves for genomic DNA and a 75 bp target.*

[h] *Estimated based on reported approximate 1 cycle delay for extraction calibration curves compared to PCR calibration curves for genomic DNA.*

Prior studies showed potential for detection of TB-specific urine cfDNA in HIV-positive and HIV-negative patients, but had variable sensitivity due to sample preparation and/or amplification methods sub-optimal for short targets (42, 52, 82). Cannas *et al.* used an in-house silica resin method (based on Promega Wizard DNA Purification Resin) reportedly designed to improve binding of short cfDNA, but did not report its analytical performance (e.g., percent recovery) (4). While the Wizard method improves upon conventional silica-based methods, our past testing revealed limited recovery (<35%) of ≤150 bp fragments (127). It was also highly dependent on urine composition and is likely to fail in samples with high pH and/or low non-target DNA concentrations (127). Labugger *et al.* and Patel *et al.* used a similar silica resin-based method, but did not specify the resin or binding conditions (5, 6). It is possible that their method suffers from similar limitations as the Wizard method used by Cannas *et al.*, although we could not experimentally verify this. They reported recoveries of approximately 30–50% and limits of detection of 3 copies/mL (75 bp target) (5) and 1.25 copies/mL (TB gDNA containing up to 42 target repeats) (6). Our sequence-specific purification method increases percent recovery by 2-fold or more and improves the limit of detection to ≤0.5 copies/mL of 50 bp cfDNA (with a positivity cutoff threshold of 0.1 copies/mL used here). Our method allows the full volume eluted from a 10 mL urine extraction to be amplified in a single PCR well, maximizing sensitivity for detection of low-concentration samples.

Past studies also enrolled few smear-negative participants, and thus have not tested the ability to detect urine cfDNA in individuals with paucibacillary TB who stand to benefit most from a non-sputum-based test. In particular, the most promising study by Cannas *et al.* included only 5% (2/43) smear-negative participants (4). Our results indicate that cfDNA detection sensitivity is higher in smear-positive compared to smear-negative individuals, so enrollment biased towards smear-positive participants may lead to an overestimation of assay sensitivity. In contrast, we demonstrated higher sensitivity while including 61% (30/49) smear-negative participants. Because Xpert MTB/RIF has reduced sensitivity (67%, pooled) in smear-negative, culture-positive sputum samples (7), our study is still limited in that it likely does not include Xpert-negative individuals with the lowest sputum bacterial loads. In a single sample available from an Xpert-negative, culture-positive participant (not included in the study described here), I was able to consistently detect TB-specific cfDNA at relatively high concentrations. Including Xpert-negative, culture-positive participants in future studies will be critical for determining the performance of the TB urine cfDNA assay for paucibacillary TB.

Although not directly tested here, an added benefit of our sequence-specific approach is that it may help ensure specificity by removing non-target DNA. Sequence-specific purification has been used to improve sensitivity of TB diagnosis from sputum by removing high concentrations of non-target DNA that can lead to downstream amplification inhibition (125), but has not been applied in urine where the primary advantage of non-target DNA removal would be to reduce the likelihood of downstream nonspecific amplification. Confirming specific amplification of short targets without the footprint for a fluorescent detection probe can be difficult, but the added layer of specificity offered by sequence-specific purification may aid in overcoming this challenge.

Comparison to existing rapid TB tests. Our urine cfDNA assay has the potential to diagnose TB in individuals who may be missed by other rapid tests (e.g., smear microscopy, urine LAM). We detected TB-specific cfDNA in the urine of all smear-positive and LAM-positive patients. Importantly, sensitivity remained high in both smear-negative (76.0%) and LAM-negative (76.5%) patients. The target product profile for a rapid biomarker-based non-sputum-based TB test outlines the optimal requirements for pulmonary TB in adults as ≥98% sensitivity for smear-positive TB, ≥68% sensitivity for smear-negative TB, and ≥80% sensitivity for HIV-associated TB (2). Our results suggest that, by employing an optimized extraction method with demonstrated high efficiency for short fragments, urine cfDNA-based assays have the potential to achieve sufficient sensitivity to meet these criteria and improve diagnostic accuracy compared to smear microscopy.

Unlike urine LAM tests, which have insufficient sensitivity, particularly in HIV-negative individuals (165), cfDNA tests have the potential to diagnose TB regardless of HIV status and CD4 count. We observed slightly higher sensitivity in HIV-positive (88.2%) compared to HIV-negative (73.3%) patients, but the difference was non-significant and the number of HIV-negative patients was small. Despite a moderate inverse correlation between CD4 count and detected TB-specific cfDNA concentration, there was no significant difference in sensitivity for HIV-positive patients with CD4 counts of ≤200 compared to >200 cells/mm^3. Although the small sample sizes led to large confidence intervals that may obscure some meaningful differences, and future studies are needed to conclusively compare cfDNA detection sensitivity across HIV-positive and HIV-negative subgroups, our results suggest that urine cfDNA is detectable across a wider patient population than urine LAM.

In contrast, the commercially-available Alere Determine TB LAM test has 42% pooled sensitivity in HIV-positive individuals, with sensitivity inversely proportional to CD4 cell count (20). The WHO only recommends its use in people living with HIV but not as a general screening test for TB (21). Ongoing

efforts aim to improve urine LAM detection sensitivity. The Fujifilm SILVAMP TB LAM test improves sensitivity relative to Alere LAM in both HIV-positive (70% vs. 42%) and HIV-negative (53% vs. 11%) individuals (22, 23). As antigen-based tests, however, LAM assays may not be able to achieve the sensitivity afforded by nucleic acid amplification tests. On the other hand, a critical advantage of LAM assays over cfDNA is their ease of sample processing. The cfDNA assay described here is not yet suitable for use in most clinically relevant settings, and will require substantial simplification in order to compete with existing rapid TB tests.

Contributions to evidence for urine cfDNA as a biomarker for TB. To date, usefulness of urine cfDNA as a biomarker for TB has been limited, in part due to inconsistent methods and results in previous TB urine cfDNA studies. Our study contributes to the evidence for urine cfDNA as a TB biomarker in two ways: 1) demonstration of the feasibility of high sensitivity and specificity with optimal pre-analytical methods and 2) development of a reliable, quantifiable method to further study TB urine cfDNA. Previous TB urine cfDNA studies have focused on measuring diagnostic accuracy and have mostly neglected to report TB urine cfDNA concentrations. Labugger *et al.* measured concentrations for a limited number of treatment-naïve individuals (n=7), which ranged from 1–41 copies/mL (median 6.5 copies/mL) (5). Here, we quantified TB-specific cfDNA to better estimate the clinical range (<1–2804 copies/mL; median 14.6 copies/mL) and have conducted analyses to determine which variables correlate with cfDNA concentration. We found that cfDNA concentration was higher in LAM-positive patients compared to LAM-negative patients, but LAM result did not affect cfDNA sensitivity. We detected cfDNA in all patients who had recently initiated TB treatment, with a higher concentration compared to treatment-naïve patients, supporting the possibility of using cfDNA for treatment monitoring as suggested previously (5). Upon initiation of a successful treatment regimen, cfDNA concentration may temporarily increase due to bactericidal activity, followed by a slow decline as the infection is cleared (5). Our study did not show a correlation between days of treatment and cfDNA concentration, but only included participants with ≤3 days of treatment and did not monitor individual participants over time.

Although cfDNA was detectable regardless of HIV status or CD4 count, the detected concentration had a moderate inverse correlation with CD4 count. We also found that cfDNA concentration had a moderate inverse correlation with days to culture positivity, suggesting that levels of excreted cfDNA may be related to bacterial burden. We observed no correlation with AFB sputum smear score or Alere urine LAM score, but the sample sizes for these analyses were small. We anticipate that our assay can be used

to continue to study TB urine cfDNA trends across subgroups and answer important unresolved questions regarding optimal sample collection techniques, processing methods, and storage conditions that have been the focus of recent work (88, 91, 92). As a caveat, the cfDNA concentrations measured by our assay may be confounded by the variable copy number of IS6110 (0–25 copies) (129) and by differences in participants' hydration status. In the future, strain typing and normalizing cfDNA concentration to urine creatinine may help better elucidate trends in cfDNA concentration.

Study limitations and future work. TB symptoms were required in addition to a positive Xpert result for the TB case definition to reduce the risk of a false-positive Xpert result, but it is possible that some could still have occurred. Using the strictest TB-positive criteria, requiring both positive Xpert and positive culture, would result in a non-significant increase in cfDNA sensitivity (to 88%). Several TB-positive, cfDNA-negative samples (false negatives) narrowly missed the cfDNA positivity cutoff, suggesting the opportunity for future improvement in clinical sensitivity. We are currently pursuing three approaches to further improve our assay's ability to detect low cfDNA concentrations (see Chapter 7): an increase in urine input volume, a reduction in PCR target length (to 25 bp using an ultrashort PCR design described in (127)) and multiplexing to target multiple genomic regions. Multiplexing will have the additional benefit of improving inclusivity by enabling detection of TB strains lacking IS6110. In subsequent clinical work, samples were stored at -80°C instead of -20°C at the collection site in South Africa (which may improve cfDNA stability) and Xpert MTB/RIF Ultra was used as the reference standard instead of Xpert MTB/RIF (which may expand the included patient population to include participants with lower bacterial loads). Limiting collection to first morning void urine or testing across multiple time points could help improve sensitivity, but would be less convenient for real-world sample collection.

This study serves as a valuable demonstration of the feasibility of our sequence-specific approach in adults with active pulmonary TB, but the sample size and scope were limited. Future studies will seek to further validate our assay and better compare sensitivity across subgroups using larger sample sizes (including more HIV-negative participants and negative controls). Subsequent work will also include expanded populations, specifically those underserved by current rapid sputum-based tests (including individuals with Xpert-negative TB, children, and patients with EPTB). TB-specific cfDNA has been detected in the urine (60) and plasma (84, 85, 164) of individuals with EPTB, including a case study of a child with tubercular otitis media (83), but there have not yet been any prospective studies in children.

Our assay also currently requires a trained user, laboratory equipment, and significant hands-on time. In the future, we aim to simplify our laboratory-based test into a format more suitable for use in resource-

limited settings, possibly by adapting technologies in development for silica-based magnetic bead purification for sequence-specific purification. Suggestions for future assay simplification are outlined in Section 8.6.4.

5.5. ROC CURVES FOR ALTERNATE POSITIVITY THRESHOLDS

The criteria used to define positive samples (≥2/3 positive replicates with ≥1 copy of cfDNA) were set prior to analysis and, while they performed well, it is worth re-examining them now that more data are available to identify optimal positivity thresholds moving forward. In particular, it would be useful to 1) confirm that specificity is robust enough and 2) identify the best threshold to use without replicate testing. There were two negative control samples with a single positive replicate. These samples were called as negative through replicate testing (which was, by design, allowed to tolerate a single positive replicate due to infrequent, late NTC amplification) but highlight the risk of false positives without replicate testing.

The optimal positivity threshold will confidently maintain specificity with minimal tradeoff in sensitivity. To help determine this threshold, Figure 5.5A shows an ROC curve treating each replicate as its own data point, rather than grouped with other replicates from the same patient (AUC = 0.92; 95% CI: 0.88–0.96). For this dataset, the optimal positivity threshold for maximizing specificity is >2.9 copies, with sensitivity and specificity of 76.9% and 100%, respectively. In contrast, for the maximal sensitivity of 85.0% (threshold >0.05 copies), specificity is reduced to 97.2%. If not conducting replicate testing, the positivity threshold should be increased from ≥1 copy to ≥2.9 copies, if not higher, to maintain specificity.

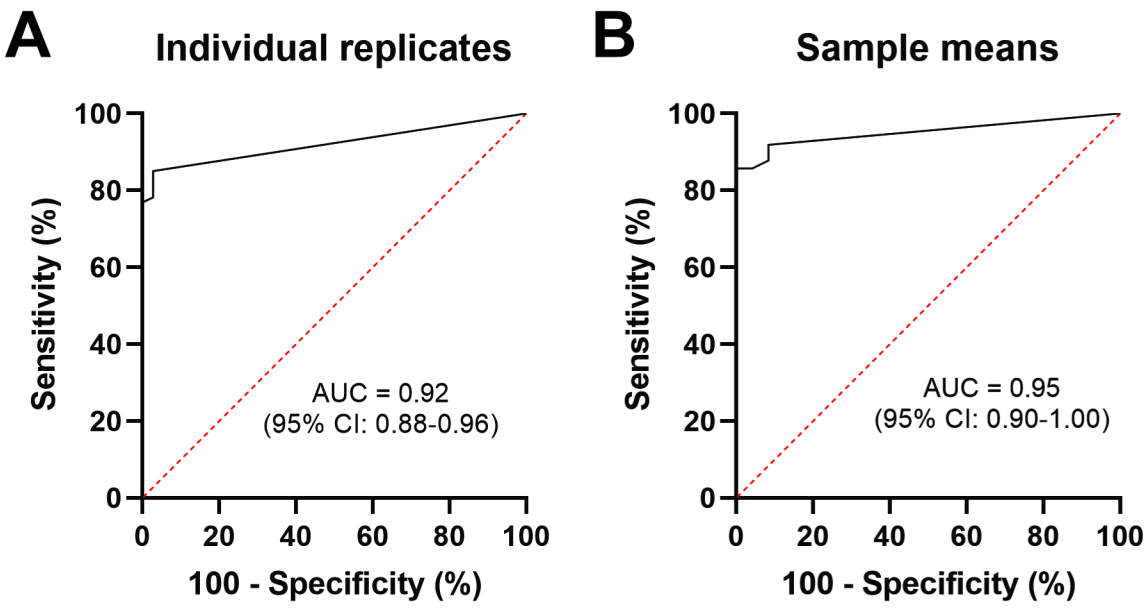


Figure 5.5: ROC curves for alternate positivity thresholds. (A) *Treating each replicate as an individual sample. This ROC curve will help determine the optimal positivity threshold for future non-replicate testing.* ***(B)*** *Analyzing sample means across n=3 replicates. This ROC curve suggests that averaging detected cfDNA concentration across replicates may perform slightly better than the current approach (calling replicates as positive/negative and requiring a certain number of positive replicates for a sample to be called as positive).*

An alternate strategy to call positive samples, in the case where replicate testing is possible, is to use the sample mean across replicates. This approach may be more effective than the current replicate-based positivity definition as it better considers the concentration detected across replicates rather than simply calling them as positive or negative. Figure 5.5B shows the ROC curve resulting from this approach (AUC = 0.95; 95% CI: 0.90–1.00). Using a positivity threshold of a sample mean of >1 copy, the sensitivity and specificity are 85.7% and 100%, respectively, which is a marginal increase in sensitivity from the current replicate-based positivity definition (83.7%). An approach based on sample means should still be paired with a T_m requirement (i.e., replicates with incorrect T_m are excluded from sample mean).

Hopefully, these ROC analyses will become irrelevant with improvements to sensitivity using the strategies described in Chapter 7, resulting in fewer samples close to the limit of detection, but they may be useful for additional clinical testing in the meantime.

5.6. Conclusions

Sequence-specific purification improves recovery of short urine cfDNA and increases the sensitivity of adult pulmonary TB diagnosis from urine cfDNA. Our assay has the highest reported accuracy of any TB urine cfDNA test to date and has the potential to enable urine-based TB diagnosis across sputum-scarce and paucibacillary populations. This study will lay the foundation for expanded clinical studies and future development of a rapid test. In addition, our work serves as a valuable contribution to the clinical evidence for urine cfDNA as a biomarker for TB. The ability to diagnose TB across key underserved populations (e.g., children, people living with HIV, individuals with EPTB) using urine samples would address an urgent need that was identified as one of the highest priority gaps in TB diagnostics (2). A sensitive urine-based test built upon the sequence-specific purification method described here could significantly contribute to improving sample availability, expanding access to rapid TB diagnosis, and controlling the TB epidemic.

5.7. Funding acknowledgements

Research reported in this chapter was funded by the Bill and Melinda Gates Foundation under award number OPP1152864 and the National Institute of Allergy and Infectious Diseases of the National Institutes of Health under award number R21AI125975. This research was funded in part by a 2018 CFAR developmental grant from the University of Washington/Fred Hutch Center for AIDS Research, an NIH-funded program under award number AI027757 which is supported by the following NIH Institutes and Centers: NIAID, NCI, NIMH, NIDA, NICHD, NHLBI, NIA, NIGMS, NIDDK. A.O. was supported by funding from the National Science Foundation Graduate Research Fellowship Program.

6. Next-generation sequencing of tuberculosis urine cell-free DNA

This chapter is adapted from

Unbiased sequencing of *Mycobacterium tuberculosis* urinary cell-free DNA reveals extremely short fragment lengths. Oreskovic A, Waalkes A, Holmes EA, Rosenthal CA, Wilson DPK, Shapiro AE, Drain PK, Lutz BR, Salipante SJ (manuscript in preparation for submission to Journal of Clinical Microbiology).

6.1. Abstract

Urine cell-free DNA (cfDNA) presents an attractive target for diagnosing pulmonary *Mycobacterium tuberculosis* (TB) infection but has not been thoroughly characterized. Here, we aimed to investigate the

size and composition of TB-derived urine cfDNA with minimal bias using next-generation DNA sequencing (NGS). To enable analysis of highly fragmented urine cfDNA, we used a combination of DNA extraction (Q sepharose) and single-stranded sequence library preparation methods demonstrated to recover short, highly degraded cfDNA fragments. We examined urine cfDNA from ten HIV-positive patients with confirmed pulmonary TB (nine of which had TB cfDNA detectable by qPCR) and two TB-negative controls. TB-derived cfDNA was identifiable by NGS from all TB-positive patients. TB urine cfDNA was significantly shorter than human urine cfDNA, with median fragment lengths of ≤19–52 bp and 42–92 bp, respectively. TB cfDNA abundance increased exponentially with decreased fragment length, with a peak fragment length of ≤19 bp in most samples. Our methodology also revealed a larger fraction of short human genomic cfDNA than previously reported, with peak fragment lengths of 29–53 bp. Urine cfDNA fragments spanned the TB genome with relative uniformity, but nucleic acids derived from multicopy elements were proportionately overrepresented, providing regions of inherent signal amplification beneficial for molecular diagnosis. This study demonstrates the potential of urine cfDNA as a diagnostic biomarker for TB and will inform improved design of TB urine cfDNA assays. Methods capable of targeting the shortest cfDNA fragments possible will be critical to maximize TB urine cfDNA detection sensitivity.

6.2. Introduction

There is a critical need for diagnostics for pulmonary *Mycobacterium tuberculosis* (TB) infection that do not rely on the collection of sputum samples, which are difficult to obtain from many patients. Even when sputum is available for testing, existing sputum-based TB assays have reduced sensitivity for diagnosis of paucibacillary, HIV-associated, pediatric, and extrapulmonary TB (7–9). Transrenal urine cell-free DNA (cfDNA) is a promising, easy-to-collect biomarker for TB with the potential to diagnose TB across these key underserved patient populations (4–6, 42, 52, 163), but it has not yet been extensively characterized or validated as a diagnostic analyte.

In particular, the fragment length distribution of transrenal microbial cfDNA, including TB cfDNA, has not yet been robustly determined. Recent work employed NGS sequencing of three TB urine cfDNA samples using conventional methods and identified the presence of short fragments (19–44 bp) but provided minimal characterization of the overall fragment length distribution (166). Moreover, conventional methods for DNA extraction and next-generation sequencing (NGS) may systematically underestimate the proportion of short urine cfDNA molecules present because they have poor retention of degraded DNA fragments. In contrast, single-stranded NGS library preparation (167, 168) improves yield of short

cfDNA fragments (<100 bp) and reveals the presence of highly degraded forms of cfDNA, including microbial cfDNA, that are less protected from nuclease digestion than human genomic cfDNA (Figure 6.1) (49, 55).

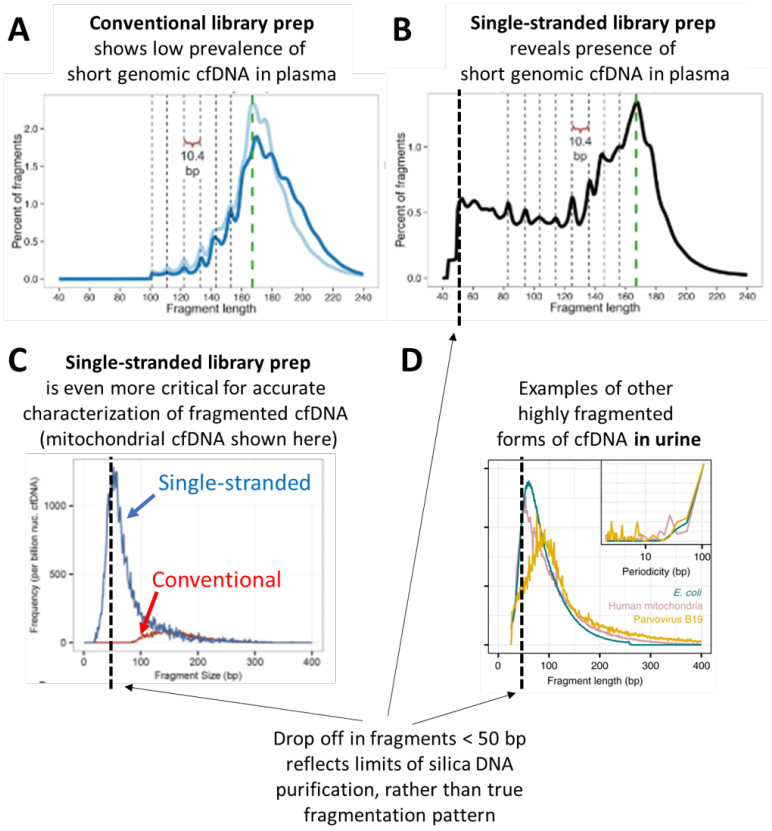


Figure 6.1: Conventional sample preparation methods (DNA extraction and library preparation) limit understanding of fragmented cfDNA. (A) *Conventional library preparation methods use blunt-end ligation of double-stranded adapters and require size selection to eliminate adapter dimers, preventing sequencing of short and/or single-stranded fragments. Plasma cfDNA is shown here, but bias against short fragments is even more dramatic for urine cfDNA, which is more degraded than plasma cfDNA.* ***(B)*** *Single-stranded library preparation retains short and single-stranded fragments, revealing previously uncharacterized, highly fragmented cfDNA in plasma (shown here) and urine (not shown).* ***(C)*** *For more degraded forms of cfDNA (such as mitochondrial cfDNA shown here, single-stranded library preparation is essential for accurate characterization of short fragments.* ***(D)*** *All cfDNA is urine is more fragmented than in plasma, but especially so for non-nucleosomal sources of cfDNA, including bacterial, mitochondrial, and viral cfDNA. The observed drop-off in fragments below approximately 50 bp is likely due to the lower size limit of silica DNA purification, not the true fragmentation pattern of cfDNA. Images reprinted and adapted from Snyder et al. (2016)* (51) *with permission from Elsevier, Burnham et al. (2016)* (49) *under a CC BY 4.0 license, and Burnham et al. (2018)* (55) *under a CC BY 4.0 license.*

In this study, we aimed to better characterize the fragment length distribution of TB-derived urine cfDNA using NGS and identify any potentially over-represented sequences suitable for targeting in diagnostic assays. In addition to using single-stranded library preparation to improve the recovery of

highly degraded DNA fragments, we selected a DNA extraction method based on Q Sepharose anion exchange resin known to increase retention of short urine cfDNA relative to other urine cfDNA extraction methods, including commonly-used commercial extraction kits (79, 127). We theorized that this combination of methods would minimize biases resulting from fragment length and enable more accurate NGS characterization of TB urine cfDNA.

6.3. METHODS

Participant enrollment and urine collection. Participants were enrolled at Edendale Hospital in Pietermaritzburg, South Africa between October 2019 and February 2021 (approved by the University of KwaZulu-Natal Biomedical Research Ethics Committee, #BE475/18). Adults (≥16 years old) with a positive admission sputum Xpert MTB/RIF Ultra (Cepheid, Sunnyvale, CA, USA) or adults with HIV, regardless of reason for admission, were recruited and provided sputum for Xpert Ultra testing. Patients with >24 hours of anti-TB treatment were excluded. All participants provided written informed consent.

Participants provided a urine sample (50–200 mL) at the time of enrollment and/or an early morning, first-void sample the morning after enrollment. The samples were not obtained mid-stream. Immediately after collection, urine was mixed in 10 mL aliquots with EDTA to a final concentration of 25 mM and Tris-HCl pH 7.5 to a final concentration of 10 mM. Urine was stored in DNA LoBind tubes (Eppendorf, Hamburg, Germany) at -80°C until shipping, shipped on dry ice to the University of Washington, and stored at -80°C until processing. Samples were de-identified prior to testing.

Clinical data, sputum testing, and urine lipoarabinomannan (LAM) testing. Clinical data were collected as follows for all participants: gender, presence of TB symptoms (cough, fever, night sweats, weight loss, loss of appetite, fatigue), TB treatment duration, HIV test result, and $CD4^+$ cell count (for participants living with HIV only). Expectorated sputum was submitted to the South African National Health Laboratory System (NHLS) for Xpert MTB/RIF Ultra testing (if not already performed) and confirmatory solid and liquid mycobacterial culture. Mycobacterial culture was performed at the NHLS Provincial TB Reference Laboratory using both Middlebrook 7H11 solid agar medium and the liquid BACTEC mycobacterial growth indicator tube (MGIT) 960 system (BD, Franklin Lakes, NJ, USA) for each sputum sample. Cultures were incubated for up to 42 days. Culture plates were read at 3 and 6 weeks, and *M. tuberculosis* was identified from solid or liquid cultures using niacin and nitrate testing. Participants were considered culture-positive with growth from either the solid or liquid culture. Urine (60 µL) was tested using Alere Determine TB LAM Ag (Abbott Laboratories, Chicago, USA). Participants were categorized as TB-positive if at least one of the Xpert MTB/RIF Ultra or mycobacterial culture were

positive. Participants were categorized as TB-negative if neither Xpert MTB/RIF Ultra or culture were positive and no clinical diagnosis of TB was assigned within 2 months of enrollment.

cfDNA extraction using Q Sepharose anion exchange resin. Urine was thawed at 37°C and centrifuged for 5 minutes at 8,000g to pellet cell debris. The cell-free urine supernatant was transferred to new 15 mL DNA LoBind tubes (Eppendorf). cfDNA was extracted using a previously-described Q Sepharose extraction protocol (79, 127), with some modifications. Specifically, we increased the spin speed of the silica column wash steps from 800g to 8,000g to avoid column clogging that occurred when processing some clinical samples, reduced the elution volume from 106 µL to 50 µL to increase the concentration of eluted cfDNA prior to library preparation, and reduced the volume of eluate per PCR well from 5 µL to 2 µL to avoid PCR inhibition. We found that these modifications had no effect on the total cfDNA yield or recovery of a spiked 50 bp positive control sequence (not shown). Each 10 mL urine sample was mixed with 300 µL Q Sepharose Fast Flow resin (GE Healthcare, Waukesha, WI) and rotated at room temperature for 30 min. The resin was pelleted for 5 minutes at 1,800g, resuspended in 1 mL low salt buffer (0.3 M LiCl, 10 mM NaOAc pH 5.5), transferred to a Mini Bio-Spin Column (Bio-Rad Laboratories, Hercules, CA), and filtered for 1 minute at 800g. The resin was washed four times with 0.5 mL low salt buffer for 30 seconds at 800g. DNA was eluted using 670 µL high salt buffer (2 M LiCl, 10 mM NaOAc pH 5.5) for 3 minutes at 800g. The eluate was mixed with 2 mL 95% ethanol and applied in 700 µL increments to a QIAquick column (Qiagen, Hilden, Germany) with 30 seconds centrifugation at 8,000g for each increment. The column was washed twice with 0.5 mL 2 M LiCl in 70% ethanol and twice with 0.5 mL 75 mM KOAc pH 5.5 in 80% ethanol for 30 seconds at 8,000g. The column was dried for 3 minutes at 20,000g and DNA was eluted in 50 µL elution buffer (Qiagen) for 2 minutes at 20,000g.

Quantification of cfDNA. The total concentration of purified cfDNA was measured using the Qubit dsDNA HS Assay Kit (Thermo Fisher Scientific, Waltham, MA, USA). The concentration of *Mycobacterium tuberculosis* (MTB) complex-specific cfDNA was measured using qPCR of a 40 bp region of the variable copy number insertion sequence IS6110. PCR was carried out in triplicate for each sample, with 2 µL eluate amplified in a 50 µL reaction containing 1.25 U OneTaq Hot Start DNA Polymerase (New England Biolabs [NEB, Ipswitch, MA, USA]), 1X OneTaq GC Reaction Buffer (NEB; 80 mM Tris-SO_4, 20 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 5% glycerol, 5% dimethyl sulfoxide, 0.06% IGEPAL CA-630, 0.05% Tween-20, pH 9.2), 0.8 mM dNTPs (NEB), 0.4X EvaGreen (Biotium, Fremont, CA, USA), 200 nM forward primer (5'-CGAACCCTGCCCAGGTCGA-3'), and 200 mM reverse primer (5'- GTA+GCAGA+CCTCACCTATGTGT-3', where "+G" and "+C" indicate locked nucleic acid bases). Primers were ordered from Integrated DNA

Technologies (Coralville, IA, USA) with HPLC purification. qPCR was carried out in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) using an initial incubation period of 94°C for 3 minutes, followed by 45 amplification cycles of 94°C for 30 seconds, 64°C for 30 seconds, and 68°C for 1 minute. Quantification cycle (C_q) values were determined using the CFX Maestro software version 1.1 (Bio-Rad Laboratories) at a threshold of 500 RFU and recovered copies were calculated using a standard curve. PCR products were confirmed by post-amplification melt curve analysis from 65°C to 95°C in 0.5°C increments every 5 seconds.

Next-generation sequencing. Sequencing libraries were prepared with 1 to 45 ng purified cfDNA using the SRSLY method as described elsewhere (168), or using the commercially available formulation of that protocol, SRSLY PicoPlus kit (Claret Biosciences), with some modifications. Specifically, to retain low molecular weight cfDNA fragments, Monarch Genomic DNA Purification Kit (NEB) was used for purification of library fragments after the phosphorylation/ligation reaction, final library purification was performed using 1.8X volumes AMPure XP beads (Beckman Coulter), and size selection of the final library was not performed. Two to 11 replicates were generated per specimen. Sequencing was carried out using a NextSeq500 (Illumina) with 150bp Paired-End chemistries.

Data analysis. Sequence reads from replicate library preparations of a given specimen were combined prior to data analysis. Reads were trimmed using the fastq-mfc algorithm of ea-utils-1.1.2.779 (169), with minimum post-trimmed sequence length set to 15 bp. The taxonomic composition of processed reads was cataloged using kraken2 (170). Individual read pairs classified as human (downsampled to 200,000 reads) or as members of the *Mycobacterium* genus were isolated for further analyses. Sequence read mapping was performed against the human genome (hg38) or the *Mycobacterium tuberculosis* H37Rv genome (GenBank Accession AL123456.3), respectively, using bwa-mem (v0.7.12) (171) with default parameters. Sequence reads with a minimum mapping quality score of 5 or greater were retained for further analysis. This approach was ultimately able to identify fragments ≥19 bp in length, equivalent to the default minimal seed length required by bwa-mem. cfDNA fragment length distributions were determined using deepTools (172), with the "distanceBetweenBins" flag set to 100.

For studies of multicopy elements IS6110 and IS1081, reads were mapped directly to those reference sequences (GenBank Accession X17348.1 and X61270.1, respectively) using bwa-mem as described above, and read counts were subsequently quantified. All statistical analysis was conducted using GraphPad Prism v8.1.2 (San Diego, CA, USA) with a significance level of 0.05.

Data availability. Reads mapping to the TB genome are available from the NCBI Sequence Read Archive (SRA) under accession PRJNA725220.

6.4. RESULTS

Q Sepharose DNA extraction and cfDNA quantification. We extracted cfDNA from the urine of 29 TB-positive and 5 TB-negative participants enrolled at Edendale Hospital in Pietermaritzburg, South Africa. We detected TB-specific cfDNA by Q Sepharose extraction and IS6110 qPCR in 14/29 (48.3%) of urine samples from TB-positive participants and 0/5 (0%) of urine samples from TB-negative participants. A summary of the detected concentrations of total cfDNA and TB-specific cfDNA is given in Table 6.1.

Table 6.1: Concentrations of total and TB-specific urine cfDNA detected after Q Sepharose extraction.[a]

		Median (IQR)	Range
Total cfDNA concentration[b]	**Eluate (ng/µL)**	5.1 (2.7–11.9)	1.1–85.6
	Urine (ng/mL)	25.5 (13.3–59.3)	5.4–428
TB-specific cfDNA concentration[c]	**Eluate (copies/µL)**	5.2 (0.8–6.5)	0.1–792
	Urine (copies/mL)	26 (4.0–32.4)	0.6–3958

[a] The detected concentrations of total and TB-specific cfDNA in each sample selected for sequencing are given in Appendix E.
[b] Measured using Qubit HS dsDNA kit.
[c] Measured by 40 bp qPCR targeting the variable copy number insertion sequence IS6110.

Urine cfDNA sequencing. We selected nine TB-positive samples with the highest concentrations of TB-specific cfDNA detectable by IS6110 qPCR, one TB-positive sample with no detectable TB-specific cfDNA by IS6110 qPCR, and two TB-negative samples for single-stranded library preparation and sequencing. Demographic and clinical data for the subset of participants whose urine samples were selected for sequencing are given in Appendix E, Table 13.1. All participants were HIV-positive with a median CD4 count of 141 cells/mm^3 (IQR: 59–516 cells/mm^3). Participants were 42% female and 58% male. 50% of TB-positive participants had a positive urine LAM result.

A total of 30 million to 113 million sequence reads were generated per specimen, resulting in 24 to 99 million reads after initial quality filtering to computationally remove self-ligated adaptor sequences (Appendix E, Table 13.2). Library complexity, the measured proportion of unique sequence fragments sequenced per library, was high for all cases, ranging from 95.2% to 100% per specimen.

Urine cfDNA taxonomic composition. We initially characterized the taxonomic composition of cfDNA reads using metagenomic analysis techniques (170). A summary is provided in Appendix E, Table 13.2 and the full output of the metagenomic analysis is available in the Supplemental Information for the paper upon which this chapter is based. For all cases, the majority of quality-filtered sequence reads (84.5–99.2%) corresponded to human nucleic acid, while the next most abundant taxa were attributable

to microorganisms comprising the normal skin or genitourinary microbiota, primarily species within *Actinobacteria*, *Proteobacteria*, and *Bacteroidetes* (173). The proportion of reads originating from bacteria of any kind averaged 1.71% per case (range 0.57% to 3.69%), with no difference in bacterial sequence load in comparing between TB-positive and TB-negative study participants (p=0.71 by 2-tailed T-test). The remainder of reads were distributed among higher level taxonomic classifications, viral and phage sequences, and unclassified reads.

Sequences putatively classified as originating from TB or human were first mapped against their respective reference genomes to confirm their identity prior to further analysis. Following this quality control step, some number of TB-derived reads were identified in all specimens from TB-positive participants (10/10 specimens), including the one patient lacking qPCR-detectable TB IS6110 cfDNA (Appendix E, Table 13.2). Significantly, no reads mapping to the TB genome were identified from either of the TB-negative individuals (0/2 specimens). An average of 2,332 reads originated from the TB genome in TB-positive patients (range 4 to 19,547 reads per case), corresponding to a proportion of 0.0000100% to 0.0201% of total reads.

Human and TB urine cfDNA fragment lengths. We next used sequencing data to explore the length of cfDNA fragments derived from the human and TB genomes (Figure 6.2).

Human cfDNA showed a relatively broad distribution of fragment lengths, with the abundance of reads inversely proportional to fragment length (Figure 6.2A). A periodicity in fragment abundance occurred at approximately 10 bp intervals, as expected based on nucleosome length and the results of previous cfDNA sequencing studies (45, 51, 54–56). The most abundant fragment length for human urine cfDNA ranged from 28–53 bp across samples (Appendix E, Table 13.3). The median fragment lengths for human urine cfDNA were 45–97 bp (Appendix E, Table 13.3).

The abundance of TB urine cfDNA similarly increased with decreasing fragment length. However, in contrast to human cfDNA, TB-derived cfDNA displayed no periodicity and showed a left-shifted distribution (Figure 6.2B). In samples with enough reads to determine the peak TB cfDNA fragment length, it ranged from 38–43 bp (Appendix E, Table 13.3). The median fragment lengths for TB-derived urine cfDNA were 39–97 bp (Appendix E, Table 13.3) and were significantly shorter than those for human cfDNA (p=0.02 by Wilcoxon matched pairs test).

Although this analysis recovers TB-derived cfDNA with high specificity, a drawback is that it preferentially identifies longer sequences, which have a correspondingly higher probability of containing sequence motifs that uniquely identify them as TB. Shorter fragments that legitimately derive from the

TB genome are more likely to share significant similarity with other species by homology or by chance alone, and will consequently be excluded.

To provide an analysis that is less biased with respect to sequence length, we next aligned all sequence reads to the TB and human reference genomes, regardless of their metagenomic classifications, and retained for analysis those that could be successfully mapped. A comparatively greater number (average 22,545; range 4 to 78,240) and proportion (average 0.027%; range 0.000010% to 0.081%) of reads matching the TB genome were recovered from TB-positive patients (Appendix E, Table 13.2). A small number of reads from TB-negative participants also mapped to the TB genome, suggesting minor, artifactual contributions of cfDNA from other organisms that have been mapped to the TB genome. Nevertheless, the proportions of TB-mapped reads from negative patients (0.0000318% and 0.000178% of reads, corresponding to read counts of 9 and 42, respectively) were three orders of magnitude less than the average for TB-positive patients, despite the two groups having comparable proportions of total bacterial cfDNA by metagenomic analysis. Moreover, we found no correlation between the proportion of reads mapping to the TB genome and the proportion of total bacterial reads cataloged by metagenomic analysis (Pearson correlation coefficient r=0.0708, p=0.83; Appendix E, Table 13.2), but did observe a significant positive correlation between the proportion of reads mapping to the TB genome and the proportion that were unambiguously classified as TB by our high specificity approach (Pearson correlation coefficient r= 0.6727, p=0.017; Appendix E, Table 13.2). Taken together, these findings indicate that the contributions of non-TB organisms to the analysis are minor, and that the reads being mapped to the TB genome are largely attributable to TB-derived cfDNA.

While the length distribution of human reads by this approach was consistent with our earlier results (Figure 6.2C), with the most abundant human cfDNA fragment length ranging from 29–53 bp across samples, cfDNA fragments mapping to the TB genome were substantially shorter than previously indicated (Figure 6.2D). The abundance of TB urine cfDNA increased exponentially with decreasing fragment size (Figure 6.2D) and showed a peak fragment length of ≤19 bp, the minimal size that is detectable by our analysis, in most samples (8/10) (Figure 6.2E). The median fragment lengths for TB-derived urine cfDNA (≤19–52 bp) remained significantly shorter than for human urine cfDNA (42–92 bp) (p=0.002 by Wilcoxon matched pairs test, Figure 6.2E).

These mapped reads were used for subsequent analyses.

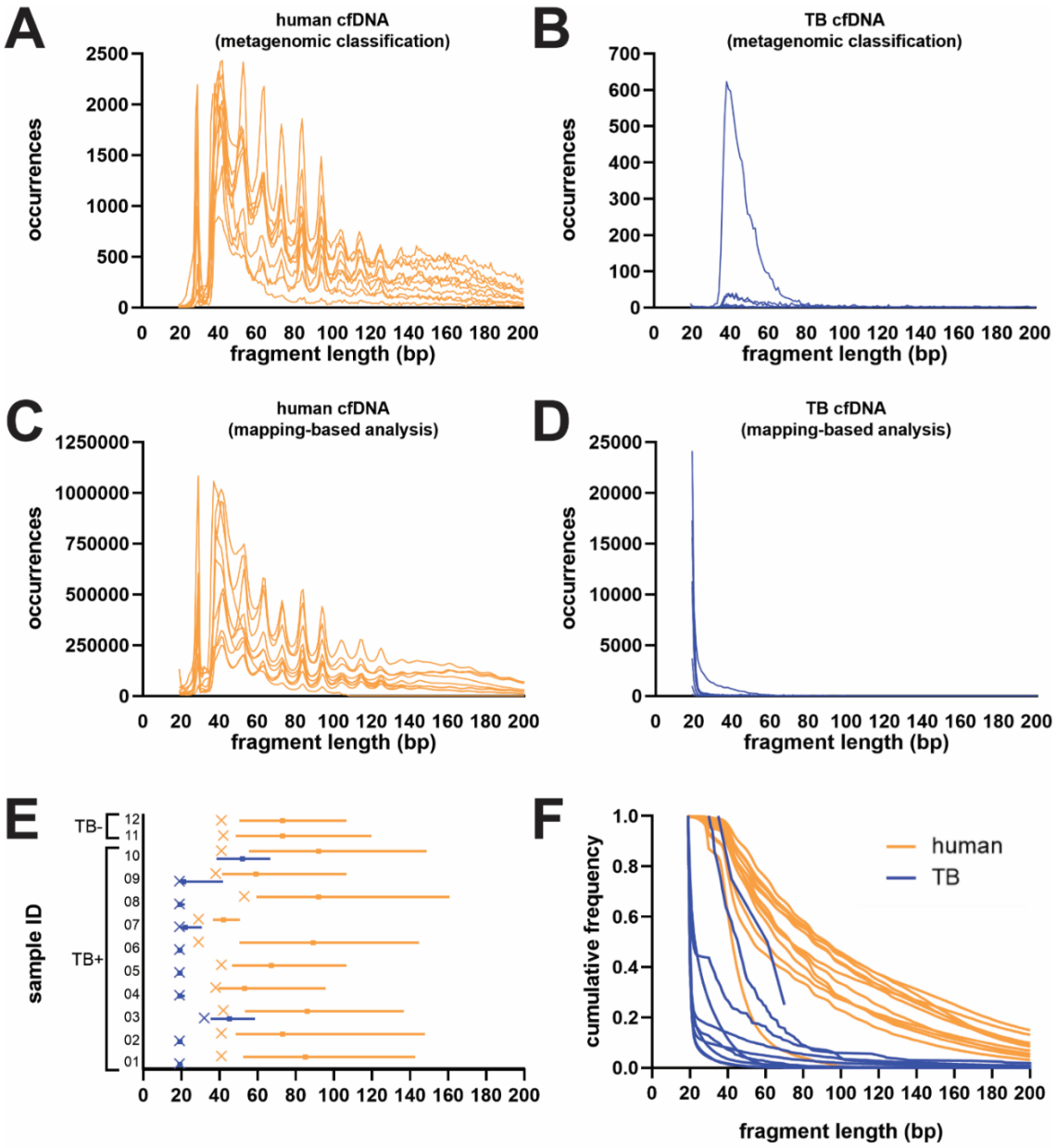


Figure 6.2: TB urine cfDNA is significantly shorter than human genomic urine cfDNA. (A) *Fragment length distributions of urine cfDNA in each sample classified as human by metagenomic analysis techniques (n=12).* ***(B)*** *Fragment length distributions of urine cfDNA in each sample classified as TB by metagenomic analysis techniques (n=10).* ***(C)*** *Fragment length distributions of urine cfDNA in each sample mapped to the human genome (n=12). Individual plots for each sample are given in Appendix E, Figure 13.1.* ***(D)*** *Fragment length distributions of urine cfDNA in each sample mapped to the TB genome (n=10). Individual plots for each sample are given in Appendix E, Figure 13.2.* ***(E)*** *Characterization of fragment length for cfDNA mapped to the TB genome (blue, n=10) and human genome (orange, n=12) in each sample. Bars indicate median fragment length and IQR. "x" indicates mode fragment length. No mode length is shown for Sample 10 because it was multimodal with a low number of reads mapped to TB. The median, IQR, and mode fragment length for each individual sample are given in Appendix E, Table 13.3.* ***(F)*** *Cumulative frequency of TB (blue, n=10) and human genomic (orange, n=12) cfDNA by fragment length in each sample.*

Distribution of TB-derived reads across the genome. In TB-positive participants, cfDNA reads mapping to the TB reference genome showed low, but relatively uniform coverage across the length of the genome (Figure 6.3A). Notably, for most samples the rRNA gene locus (positions 1471846 to 1477013 bp), evidenced increased read coverage relative to the rest of the TB genome, despite there being a single copy of this locus carried by TB (174). As rRNA encodes an essential gene that is highly conserved across bacterial taxa (175), these data suggest that short reads derived from other organisms present in patient specimens may infrequently map to the TB genome at specific sequence contexts.

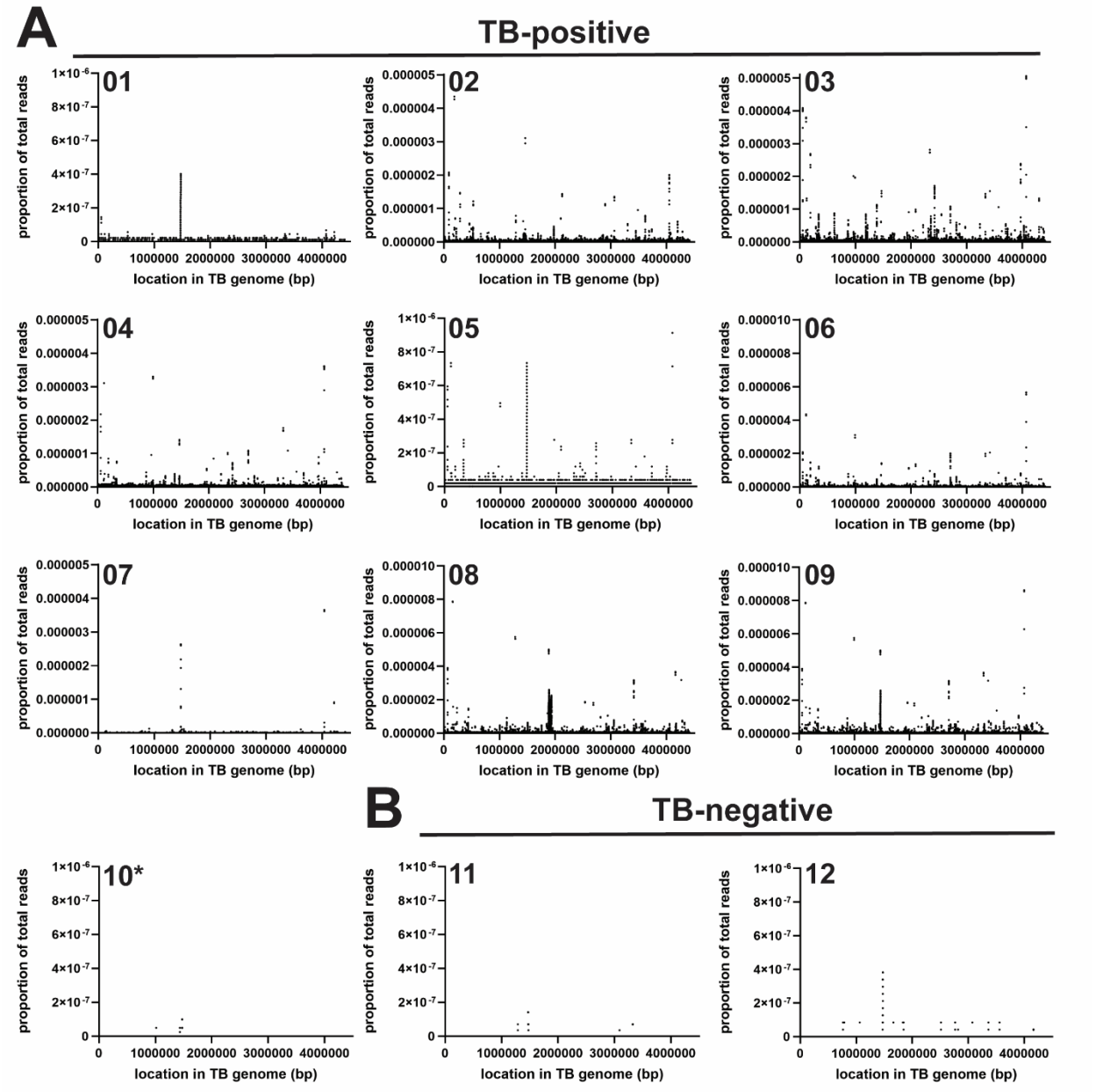


Figure 6.3: Coverage of the TB genome in urine cfDNA. (A) *Density of reads mapped to the TB genome in ten samples from TB-positive participants. *Sample 10 had no TB cfDNA detectable by IS6110 qPCR, but TB-specific cfDNA was detectable by sequencing and confirmed by metagenomic classification analysis (kraken2).* ***(B)*** *Density of reads mapped to the TB genome in two samples from TB-negative participants.*

Multicopy elements in the TB genome as diagnostic targets for urine cfDNA. Species-specific multicopy elements are attractive targets for diagnostic testing because they provide both specificity and an inherent level of signal amplification. To evaluate the potential of two known multicopy genomic elements as potential urine cfDNA diagnostic targets, we analyzed the relative abundance of reads mapping to two insertion sequences (IS6110 and IS1081) present in the TB genome (128, 176).

cfDNA fragments derived from IS6110 and IS1081 cfDNA were detected by NGS in 6/9 and 5/9 specimens with TB cfDNA detectable by IS6110 qPCR, respectively (Table 6.2). In cases where these sequences were identified, they were found with greater abundance than reads from other regions of the TB genome. The average fold overrepresentation of IS6110 relative to the average sequencing depth for the remainder of the TB genome was 8.1 (range 1.3–17.1), whereas that observed for IS1081 was 4.6 (range 1.7–8.5). These values correspond roughly to the expected count of each element per genome. IS6110 is present at variable copy number (0–25 copies) across MTB complex strains (129), while IS1081 is present at a more stable copy number (5–6 copies) (177). We were not able to examine reads mapping to a third multicopy element used in TB studies, the direct repeat (DR) region (178), given the repetitive nature of the element and the short length of its constituent repeat sequence (36 bp, present at an estimated 14–63 copies per genome), which prevented reliable read mapping.

Table 6.2: Relative abundance of multicopy elements IS6110 and IS1081 in urine cfDNA.

Sample ID	TB status	cfDNA status (IS6110 qPCR)	IS6110 fold overrepresentation[a]	IS1081 fold overrepresentation[a]
01	Positive	Positive	0	0
02	Positive	Positive	0	0
03	Positive	Positive	17.0	2.52
04	Positive	Positive	1.27	0
05	Positive	Positive	3.30	6.78
06	Positive	Positive	3.68	3.62
07	Positive	Positive	17.1	8.48
08	Positive	Positive	6.34	1.72
09	Positive	Positive	0	0
10	Positive	Negative	0	0
11	Negative	Negative	0	0
12	Negative	Negative	0	0

[a] Measured as average sequencing depth at the target region normalized to the average sequencing depth across the remainder of the TB genome.

6.5. Discussion

In this study, we present an in-depth, unbiased characterization of TB urine cfDNA using NGS. To most comprehensively characterize the range of fragments occurring in this analyte, we selected DNA extraction and sequencing library preparation methods specifically for their demonstrated effectiveness for short DNA fragments. We have previously shown in a comparison of urine cfDNA purification methods that Q Sepharose extraction, which pre-concentrates urine cfDNA using anion exchange resin prior to desalting on a silica spin column (79), has a high rate of recovery (>70%) for spiked DNA down to at least 40 bp in length (127). Recovery by that method is reduced to <10%, but is still measurable, for DNA as short as 25 bp in length (127). Similarly, single-stranded NGS library preparation (167, 168) has been shown to improve recovery of <100 bp cfDNA, having a lower reported limit of 40–60 bp (49). In this application, we have further extended the lower range of detection for this approach by retaining all library fragments generated, which increased our sensitivity for low molecular weight DNA fragments at the expense of sequencing an increased proportion of synthetic, noncontributory fragments resulting from self-ligated sequencing adaptor molecules (measured at 9% to 20% of total reads generated per specimen).

Previous NGS studies that characterized the fragment length distribution of human genomic cfDNA have reported peak fragment lengths of approximately 50–100 bp (54–56). In contrast, our study demonstrably improved recovery of short cfDNA fragments and revealed a previously undetectable fraction of human genomic cfDNA in urine having a peak fragment length of 29–53 bp, which was observed in all samples. The differences between our methodology and protocols employed previously were most noticeable for the shortest fragments, with representation of <50 bp fragments in our study dramatically increased relative to earlier work that did not use single-stranded library preparation methods (45, 54, 56) or which used single-stranded library preparation in conjunction with a DNA extraction method less able to efficiently recover short DNA fragments (Qiagen Circulating Nucleic Acid Kit) (55).

Our results provide evidence that TB cfDNA in urine is extensively fragmented, significantly more so than human genomic cfDNA (Figure 6.2). Consequently, minimizing the length of sequences targeted for TB urine cfDNA diagnostic assays will be critical to maximizing their sensitivity. Decreasing the minimum detectable target length improves detection sensitivity for fragmented cfDNA (79–81) and has been a priority during the recent development of TB urine cfDNA assays (4–6, 163). Previously reported TB urine cfDNA assays targeted, at the shortest, amplicons of 38–40 bp (5, 6, 163). Decreasing PCR

amplicon length from 49 bp to 39 bp resulted in greater than 10-fold increase in detected TB cfDNA (81). Until now, the extent to which further decreases in target length may improve sensitivity has been unclear. Our results suggest that even small, incremental decreases in target length may have a disproportionate impact on detection of TB urine cfDNA, which is highly fragmented to a peak size of ≤19 bp and increases in abundance exponentially as fragment size decreases. To improve the clinical sensitivity of TB urine cfDNA assays, both sample preparation and amplification methods having maximal efficiency for very short fragments are needed. For example, recent work in our lab demonstrated that sequence-specific purification improves recovery of short cfDNA relative to conventional silica-based extraction and increases the clinical sensitivity of TB diagnosis from urine cfDNA (162, 163). Ultrashort PCR using a stem-loop primer may be an attractive strategy for amplification of fragments too short for conventional PCR (79). Our results, in concert with those of a previous study (166), suggest that targeting multicopy genomic elements (*e.g.,* IS6110, IS1081) is likely a more promising strategy than identification of *de novo* cfDNA targets.

Our study has several limitations. First, we sequenced cfDNA only from patients living with HIV. Detection sensitivity for TB urine cfDNA is similar in HIV-positive and HIV-negative participants (6, 163), but it remains unclear if there may be differences in cfDNA fragmentation patterns across these two populations. Second, owing to the requirements for high sequencing depths and attendant sequencing costs, the number of specimens analyzed in this study are necessarily limited. Third, despite the improvements in short cfDNA fragment recovery using a combination of Q Sepharose DNA extraction and single-stranded library preparation, our methods are unable to reliably interrogate the shortest cfDNA fragments. We expect that efficiency of fragment recovery begins to decrease below 40 bp (127), and due to the nature of sequence read mapping algorithms, it is not possible to reliably map the origin or sequence reads below a specified seed length (here, 19 bp). Moreover, the shorter the fragment length, the less probable it is that the read will map confidently to its target (179). Considering these limitations, the true frequency of cfDNA less than 40 bp in length, whether originating from human or TB, is likely even greater than registered by our analysis. The fragment length distribution of cfDNA should be interpreted with this in mind. Fourth, many of the cfDNA molecules recoverable by our methods are so short that they cannot be uniquely classified as belonging to TB. As a consequence, we cannot demonstrate directly that all of the smallest read fragments we mapped to TB derive from derive from that organism, although accessory evidence is consistent with that conclusion.

6.6. Conclusions

Accurate characterization of urine cfDNA using NGS provides critical insight into its validation as a biomarker for TB. Our findings, in particular the discovery that TB cfDNA is substantially shorter than human genomic cfDNA, will help inform the development of improved assays for TB diagnosis from urine cfDNA. The large potential sensitivity benefit to be gained by targeting <40 bp TB cfDNA motivates continued prioritization of both sample preparation and amplification methods designed for short fragments, although this will need to be balanced against reduced specificity accompanied by interrogating shorter nucleotide fragments. A sensitive molecular assay targeting urine cfDNA, rather than sputum, would considerably contribute to improving sample accessibility and diagnostic yield and has the potential to advance availability of rapid TB diagnostics across underserved patient populations. In addition, the combination of Q Sepharose DNA extraction and single-stranded library preparation will likely prove useful for other applications and contexts where the analysis of highly fragmented forms of DNA is necessary.

6.7. Addendum: Comparison of Q Sepharose and sequence-specific purification for extraction of TB urine cfDNA

While not a primary objective, this study also gave us the opportunity to preliminarily compare the performance of Q Sepharose extraction and sequence-specific purification for detection of TB cfDNA in clinical urine samples. These two methods were identified as the most promising for short urine cfDNA in the analytical comparison described in Chapter 4. While sequence-specific purification performed best overall for the shortest fragments (<40 bp), Q Sepharose has the advantage of purifying total cfDNA. This is necessary for NGS and is why we used Q Sepharose rather than our in house method for cfDNA extraction in this study. Q Sepharose has not been used previously for detection of TB cfDNA.

Diagnostic accuracy. Q Sepharose detected TB cfDNA with 48.3% sensitivity (95% CI: 31.4–65.6%) and 100% specificity (95% CI: 56.6–100%) (n=34). In the study described in Chapter 4.8, sequence-specific purification detected TB cfDNA with 83.7% sensitivity (95% CI: 71.0–91.5%) and 100% specificity (95% CI: 86.2–100%) (n=74). The sensitivity of our in-house TB urine cfDNA assay using sequence-specific purification was significantly higher than when using Q Sepharose (p=0.002, Fisher's exact test). The same PCR assay was used for both studies.

There are a few caveats to this analysis. First, while the samples came from the same patient population, different samples were used to test each method. The prevalence of TB cfDNA may have differed

between the two sample pools. Second, to conserve volume for NGS and reduce the impact of PCR inhibition on quantification, I reduced the eluate volume in PCR for Q Sepharose from 5 µL (original protocol) to only 2 µL for Q Sepharose, representing 10% and 4% of the original urine volume, respectively. It is possible that amplifying a larger eluate volume would increase sensitivity. The benefit of sequence-specific purification over Q Sepharose is likely a combination of factors: improved recovery of short cfDNA, higher concentration factor (by >2-fold), and ability to amplify a larger effective fraction of the sample without complete inhibition of PCR (equivalent of 10 mL urine versus 0.4–1 mL urine).

Detected cfDNA concentration. The concentrations of 40 bp IS6110 cfDNA detected using Q Sepharose and sequence-specific purification were similar, but I calculated descriptive statistics based on only cfDNA-positive samples, so the concentrations detected by sequence-specific purification may be skewed lower because the method is better able to detect low concentration samples that Q Sepharose is likely to miss. Q Sepharose detected a median of 26 copies/mL (range: 4.0–32.4 copies/mL) and sequence-specific purification detected a median of 15 copies/mL (range 0.14–2804 copies/mL). In pooled urine from TB-positive patients, sequence-specific purification detected 8.2-fold more 40 bp IS6110 cfDNA than Q Sepharose (4012 vs 488 dsDNA copies).

Additional observations. In clinical urine samples, the Q Sepharose protocol led to difficulty processing samples and PCR inhibition that was not observed in analytical characterization experiments in Chapter 4. Specifically, many samples resulted in clogging of the silica spin column, despite pre-purification of cfDNA on Q Sepharose resin. To mitigate this problem, I increased the centrifugation speed from 800g to 8,000g for these steps, which had no measured effect on total cfDNA yield or the percent recovery of a spiked 50 bp target sequence. This change eliminated clogging for most samples, but a few still flowed slowly through the silica spin column and required additional centrifugation time. In samples from hospital inpatients, Q Sepharose also resulted in consistent PCR inhibition that was not observed in healthy control samples. This may be in part due to a reduction in elution volume (from 106 µL to 50 µL) made to better concentrate cfDNA for downstream library preparation. Reducing the eluate volume used in PCR (from 5 µL to 2 µL) helped mitigate but did not eliminate inhibition.

Conclusions. Sequence-specific purification remains superior to Q Sepharose for detection of TB urine cfDNA, and likely for other cfDNA applications requiring detection of specific target sequences. However, Q Sepharose performed well for isolation of total cfDNA prior to sequencing, which is not possible using sequence-specific purification. Both protocols require many manual steps, so neither provides an advantage for ease of use.

6.8. Funding acknowledgements

Research reported in this chapter was funded by the Brotman Baty Institute Catalytic Collaborations Program and the Bill and Melinda Gates Foundation under award number OPP1213054. A.O. was supported by funding from the National Science Foundation Graduate Research Fellowship Program.

7. Strategies to increase sensitivity: Multiplexing, ultrashort PCR, and other next steps towards continued assay development

7.1. Abstract

While the performance of our TB urine cfDNA assay in the clinical study described in Chapter 4.8 was promising, the concentration of TB cfDNA detected in many samples was low. There is an opportunity to increase clinical sensitivity and robustness of the assay by improving the detection of urine samples with low concentrations of TB-specific cfDNA. In this chapter, I describe several strategies that may help achieve this goal. The rationale for two of these strategies – multiplexing and ultrashort PCR – is to expand the detectable fraction of TB-specific cfDNA beyond the current 40 bp IS6110 target. I have developed preliminary designs for three new targets: the direct repeat (DR) region and two different subregions of the insertion sequence IS1081. I have also designed an ultrashort PCR capable of amplifying a 25 bp sequence within the current IS6110 target. The third strategy is to increase the input sample volume, which would proportionally increase the cfDNA copy number available for detection. The final strategy is to eliminate the effect of sample variability on assay performance. A subset of clinical samples, most often those with proteinuria or blood, led to difficulty during the magnetic bead handling steps (bead loss during washes and/or bead transfer to PCR during elution). Minimizing the effects of sample variability by addition of a proteinase K digestion step may improve the ease of sample processing, make the assay more amenable to adoption by new users, and possibly improve sensitivity in previously challenging samples by reducing bead loss and PCR inhibition. I have taken the first steps towards implantation of each of these strategies to identify which are most promising for continued development. In addition to increasing clinical sensitivity, improving the ability to reliably detect low concentration samples will ideally also avoid the need for replicate testing and thus increase the throughput of future clinical studies.

7.2. Opportunity for improved detection of low concentration samples

Despite the strong analytical performance of the TB urine cfDNA assay (limit of detection <0.5 copies/mL), the low concentration of TB-specific cfDNA in detected in urine (median 14.6 copies/mL, range 0.1 – 2804 copies/mL) may be a limiting factor for clinical sensitivity. At low target concentrations, PCR is limited by the Poisson distribution of target molecules in the sample. The Poisson distribution is given by the equation $P(x;\lambda) = \frac{e^{-\lambda}\lambda^{x}}{x!}$, where, for this application, λ represents the average concentration of target molecules across the entire urine sample and x represents the concentration of target molecules in an individual 10 mL sample aliquot to be analyzed. As an example, if a patient's urine contains an average of 0.1 copies/mL (λ = 1 copy per 10 mL) of the target cfDNA sequence, the probabilities of a 10 mL urine aliquot containing x = 0, 1, 2, or 3 copies of the target cfDNA sequence are 0.37, 0.37, 0.18, and 0.06, respectively. In other words, there is a 37% chance that a given aliquot will contain no detectable target cfDNA.

For the clinical study described in Chapter 4.8, we attempted to compensate for the expected occasional PCR drop-out of low concentration samples by testing urine from each participant in triplicate and requiring ≥2/3 replicates positive for a sample to be called as positive. The results of this study are shown again in Figure 7.1, this time emphasizing samples near the cutoff threshold (hollow dots). Even with triplicate testing, there were several potential "near miss" samples (hollow cyan dots; 2/3 replicates positive) and potential "barely missed" samples (hollow red dots; 1/3 replicates positive). Using our sample positivity criteria, set prior to testing, we calculated an assay sensitivity of 84% (n=41/49). If we were able to successfully detect the "barely missed" samples, which may be possible if we can improve detection of low concentration samples, the sensitivity would increase to 92% (n=45/49). On the other hand, if we did not detect the "near miss" samples, which is a possibility if the current assay were to be used without replicate testing, the sensitivity would decrease to 73% (n=36/49). Ideally, in the future we will be able to test only a single replicate per participant while more confidently detecting low concentration samples, which would allow for clinical studies with larger sample sizes and be more relevant for eventual implementation as a real-world diagnostic test.

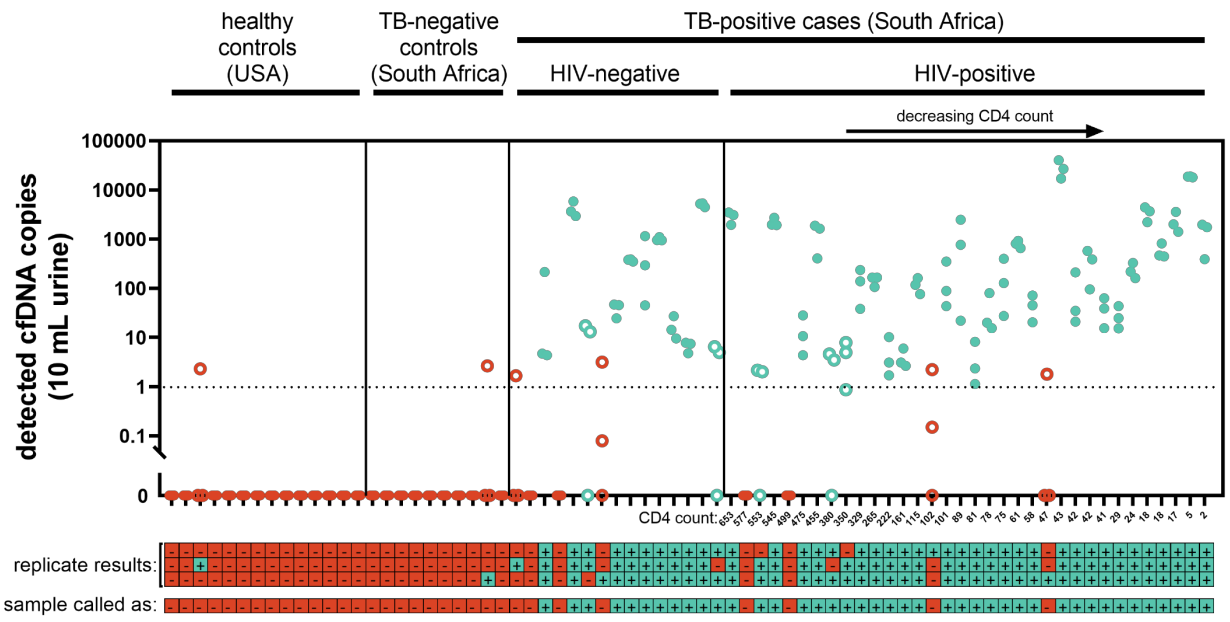


Figure 7.1: Visualization of samples near the limit of detection of current TB urine cfDNA assay. *Solid cyan dots represent samples confidently called as positive (3/3 replicates positive; n=36/49 TB-positive samples). Hollow cyan dots represent potential "near miss" samples called as positive (2/3 replicates positive; n=5/39 TB-positive samples). Hollow red dots represent potential "barely missed" samples called as negative (1/3 replicates positive; n=4/39 TB-positive samples). Solid red dots indicate samples confidently called as negative (0/3 replicates positive; n=4/39 TB-positive samples).*

A critical distinction is that the clinical sensitivity is limited not by the total concentration of TB cfDNA in urine, but rather by the low copy number of specific detectable TB cfDNA target molecules in urine. The current assay targets a single sequence: a 40 bp region of IS6110. Even for samples with low concentrations of TB urine cfDNA, there is undoubtedly more TB-derived cfDNA present in the urine, but it is not detectable using our current assay design. By expanding the fraction of detectable TB cfDNA fragments, we may be able to improve detection of low concentration samples, increase clinical sensitivity, and avoid the need for replicate testing. If the fraction of detectable fragments can be scaled up dramatically, this may also allow for a reduction in input sample volume, which would greatly ease future simplification of the assay to an automated format more suitable for use in low-resource settings. In the following sections, I describe the first steps towards implementation of several strategies to improve detection of low concentration samples, with the goal of identifying which strategies are most promising for future assay development.

7.3. First steps towards multiplexed assay

The first strategy to increase sensitivity is the development of a multiplexed assay. Multiplexed capture – whether targeting multiple subregions within a single genomic region or sequences across multiple genomic regions – will increase the number of TB cfDNA fragments available for amplification (Figure 7.2). The advantage of multiplexed capture for detecting low concentration samples could take one of two forms. Multiple targets could be amplified in a single reaction using either a nonspecific intercalating dye (where the signal across targets would be cumulative) or sequence-specific fluorescent detection probes (where the detection would also be multiplexed, with unique signals for each target). In both cases, multiplexing should help improve detection of low concentration samples that are currently near the limit of detection of PCR by increasing the likelihood that at least one target amplifies successfully. Multiplexing to include other genomic regions should also improve inclusivity by enabling detection of MTB strains lacking IS6110, the prevalence of which varies geographically (129).

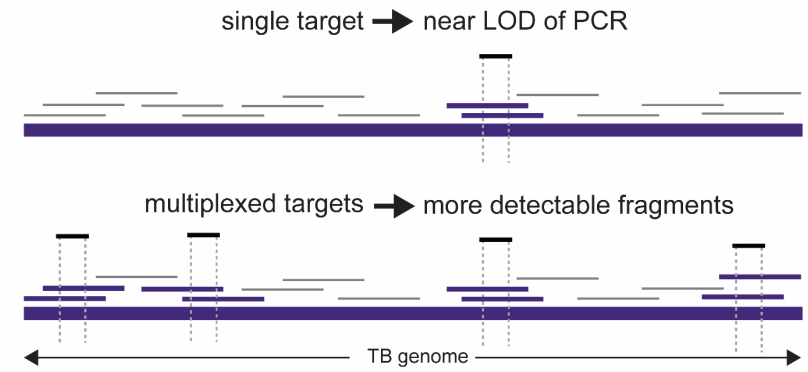


Figure 7.2: Rationale for multiplexed capture to increase sensitivity. *Detectable fragments are shown in purple.*

7.3.1. *Demonstration of multiplexing feasibility*

This subsection is adapted from

Oreskovic A, Lutz BR. Ultrasensitive hybridization capture: reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. PLOS ONE 16(2): e0247851 (2021). https://doi.org/10.1371/journal.pone.0247851.

To demonstrate the potential multiplexing capability of hybridization capture and estimate the maximum number of probes (i.e., the minimum concentration per probe) that maintains high recovery, I varied the ratio of target to non-target probes while keeping the total probe concentration constant (Figure 7.3A). I found that recovery remained the same for conditions representing up to four probes

(two probe sets targeting two dsDNA targets or four individual probes targeting four ssDNA targets). Recovery dropped moderately to 63% of the original for 10 probes (targeting five dsDNA targets or 10 ssDNA targets) and was substantially reduced to 26% of the original for 20 probes (targeting 10 dsDNA targets or 20 dsDNA targets). Increasing the hybridization time from 30 minutes to 60 minutes did not rescue recovery for the equivalent of 10 probes (Figure 7.3B). Small-scale multiplexing is possible using the current protocol, and larger-scale multiplexing may be improved through additional optimization specifically for multiplexed capture, such as a higher hybridization temperature or increased bead volume. Unique bead pools for each capture probe (rather than multiple probes on all beads) could also be explored.

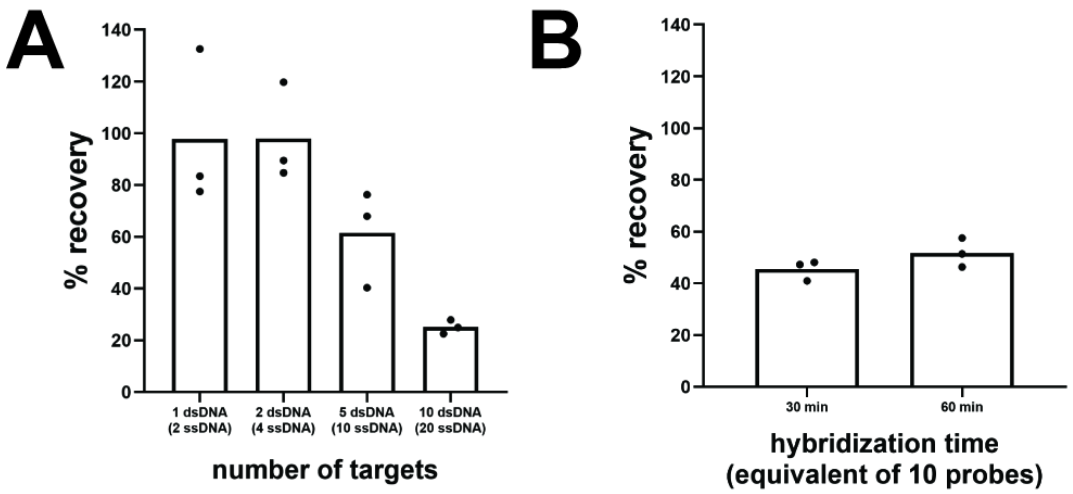


Figure 7.3: Preliminary demonstration of multiplexing capability. (A) *Without additional assay optimization, multiplexed capture of up to 5 dsDNA targets or 10 ssDNA targets is possible with only moderate reduction in recovery compared to single-plex capture. To represent conditions expected during multiplexed capture, the ratio of target-specific probes to non-target probes was varied (100:0, 50:50, 20:80, 10:90) while keeping the total probe concentration constant (50 pmol per 50 µL beads) (n=3 technical replicates per condition; 10^3 copies 50 bp dsDNA input).* ***(B)*** *Increasing the hybridization time does not improve recovery when using the equivalent of 10 probes.*

7.3.2. *Design of assays for new targets*

Our current assay targets the insertion sequence IS6110 (128, 129), which is present at variable copy number (1–28 inserts) in approximately 99% of MTB complex strains (unpublished data courtesy of Dr. Norman D. Brault). For the design of new assays, I prioritized other multiple copy genomic elements to boost the quantity of cfDNA available for detection. Insertion sequence IS1081 is present in all MTB complex strains at a more stable copy number (5–6 inserts) (176, 177). It has been previously used for detection of TB cfDNA in plasma (164). The direct repeat (DR) region is present across all MTB strains

and contains an estimated 14–63 (average 38) copies of a short repeat sequence (36 bp) separated by short (36–42 bp) spacers (178). It has been used previously for detection of TB cfDNA in urine (6).

I designed new PCR primers and capture probes targeting IS1081 and the DR region (Table 7.1) using the design criteria outlined in Section 3.8. For IS1081, I designed two assays targeting different subregions, referred to as 5' and 3' relative to their location in the insert sequence (GenBank Accession #X61270.1). Sequence-specific purification was carried out as described in Section 3.4. qPCR was carried out in a CFX96 Touch Real-Time PCR Detection System (Bio-Rad) using an initial incubation period of 94°C for 3 minutes, followed by 45 amplification cycles of 94°C for 30 seconds, annealing temperature (T_a, Table 7.1) for 30 seconds, and 68°C for 1 minute.

Table 7.1: Multiple-copy genomic elements for multiplexed assay development

Genomic element	Copies per genome	Prevalence (% of MTB complex strains)	Capture probe sequences	Primer sequences	Positive control sequence	PCR T_a
IS6110	1–28	99%		See Chapter 3		64°C
IS1081 (3')	5–6	100%	/52-Bio/AAAAAAAAAAAAAAAAAAAA TCGAGTACCCGATCATA/3SpC3	CCGAATCAGTTGTTGCCCA	CCGAATCAGTTGTTGCCC A**ACACCATTCA**ATATGA TCGGGTACTCGACGC	57°C
			/52-Bio/AAAAAAAAAAAAAAAAAAAA AATCAGTTGTTGCCCA/3SpC3	GCGTCGAGTACCCGATCATA		
IS1081 (5')			/52-Bio/AAAAAAAAAAAAAAAAAAAA ATTGACTTTTGCTGGTCG/3SpC3	TCGCGTGATCCTTCGAAACG	TCGCGTGATCCTTCGAA ACG**ACACCATTCA**CGAC CAGCAAAAGTCAATCGA	58°C
			/52-Bio/AAAAAAAAAAAAAAAAAAAA CGTGATCCTTCGAAACG/3SpC3	TCGATTGACTTTTGCTGGTCG		
Direct Repeat (DR) region	14–63	100%	/52-Bio/AAAAAAAAAAAAAAAAAAAA TCAGACCCAAAACCCC/3SpC3	GTTTCCGTCCCCTCTCG	GTTTCCGTCCCCTCTCG**A CACCATTCA**GGGTTTTG GGTCTGACGAC	56°C
			/52-Bio/AAAAAAAAAAAAAAAAAAAA TCCGTCCCCTCTCGG/3SpC3	TCGTCAGACCCAAAACCC		

/52-Bio/ indicates dual biotin modification; /3SpC3/ indicates carbon spacer. Target-specific probe binding sequences are underlined. A synthetic spacer region introduced to differentiate the positive controls from the endogenous Mycobacterium tuberculosis complex-specific target sequences is bolded.

7.3.3. *Preliminary testing of assays for new targets*

The performance of the IS1081-5', IS1081-3', and DR assays is shown in comparison to the original IS6110 assay in Figure 7.4. The new assays averaged 109%, 63%, and 78% recovery, respectively, of

positive control DNA spiked into urine (Figure 7.4A). In pooled urine from eight TB-positive patients, an average of 8047 copies IS6110 cfDNA, 3691 copies IS1081-5' cfDNA, 1162 copies IS1081-3' cfDNA, and 124 copies DR cfDNA were detected (Figure 7.4B). Although neither IS1081 target outperformed the original IS6110 assay in terms of detected cfDNA concentration in pooled urine, IS1081 detection may be more consistent across MTB strains given its more stable copy number. It would not be surprising if the benefit of including IS1081 as a diagnostic target is more apparent when testing individual TB-positive samples. Furthermore, it should be noted that, because the samples pooled for this experiment were from patients infected with unknown MTB strains (and thus unknown IS6110 copy number), a one-to-one comparison of IS6110 vs IS1081 concentration is not appropriate and should not be interpreted as "superiority" of the IS6110 assay over IS1081. Instead, these results demonstrate that IS1081 is an attractive target for inclusion in a future multiplexed assay. Additional testing of IS6110 and IS1081 detection rates and cfDNA concentrations should be conducted across a panel of TB-positive samples.

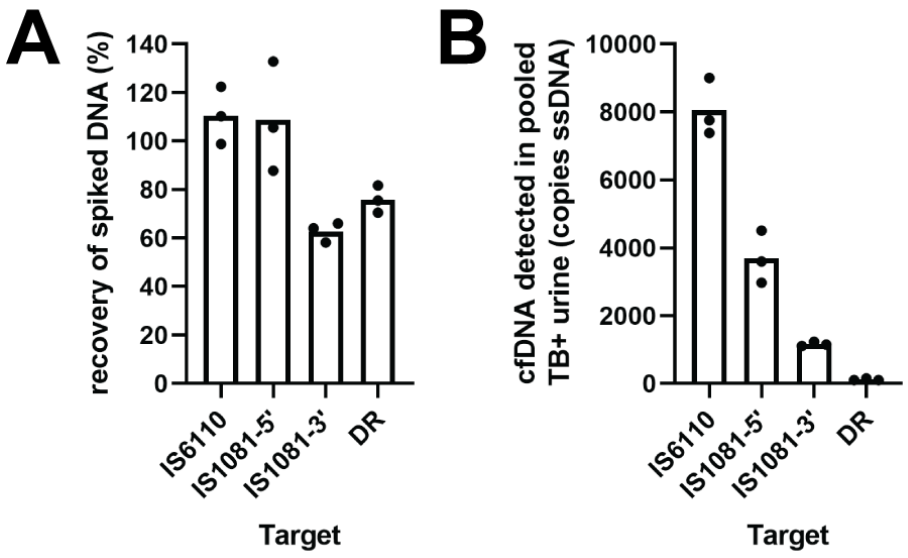


Figure 7.4: Performance of assays for new targets. (A) *Percent recovery of 1000 copies spiked dsDNA positive control for each target using sequence-specific purification (mean, n=3).* ***(B)*** *Detected copies of IS6110, IS1081, and DR cfDNA in pooled urine from eight TB-positive individuals (mean, n=3).*

Differences in detected IS1081 cfDNA concentration using the IS1081-5' and IS1081-3' assays can likely be attributed to differences in percent recovery as well as 3' truncation of some IS1081 inserts. The IS1081-3' assay had noticeably lower recovery than that of the IS1081-5' assay, for unknown reasons. I designed the IS1081-3' assay to target the same amplification region as the IS1081 target included in the Xpert MTB/RIF Ultra test (180), but this region is truncated in some inserts (for example, 1/6 inserts in

the commonly-used reference strain H37Rv). The IS1081-5’ assay, on the other hand, targets a region conserved across IS1081 inserts that has been used previously by several groups (181–183).

Compared to IS6110 and IS1081, the DR assay did not perform as well, with detected cfDNA concentration at least an order of magnitude lower. In TB-positive urine, the DR amplicon had higher than expected melting temperature (82–83°C vs 73.5°C for the positive control), likely due to amplification across multiple repeat units that are in close proximity within the DR region.

All three new PCRs could benefit from additional specificity optimization. PCR targeting the 5’ subregion of IS6110 had amplification in water NTCs (>40 cycles) and negative control urine (equivalent of ~10 copies). PCR targeting the 3’ subregion of IS6110 had occasional late amplification in water NTCs (>45 cycles) but not negative control urine. DR PCR had no amplification of water NTCs or negative control urine up to 45 cycles but amplification (>40 cycles) of 10 ng spiked human gDNA. For the sake of expediency and because these assays will likely need to be updated for inclusion in a multiplexed PCR assay, I tolerated sub-optimal specificity in these preliminary experiments. Nonspecific amplification was insignificant compared to detected cfDNA concentration and did not affect quantification.

7.3.4. *Conclusions and next steps towards multiplexed assay*

This work demonstrated the relative ease with which sequence-specific purification assays for new targets can be developed. Multicopy elements IS6110 and IS1081 but not the DR region should be prioritized for inclusion in a multiplexed assay. Assays targeting additional subregions of IS6110 (not developed here) should also be considered. Future work should evaluate potential new assay designs across individual samples, improve PCR specificity, optimize for higher-plex capture (if desired), and develop a multiplexed assay. Ideally, capture and amplification of several targets can be combined into a single assay, either with or without multiplexed detection. While multiplexed capture should be straightforward, multiplexed PCR of short targets without the footprint for sequence-specific detection probes may be more challenging. Strategies for short-target multiplexed PCR include stem-loop primers (similar to the 25 bp ultrashort PCR design in Section 7.4, but longer) to create more room for a sequence-specific detection probe (79) and double-stranded primers that enable detection via separation of fluorophore/quencher pairs (6, 184). Alternatively, a simpler approach would be to use a nonspecific DNA binding dye to detect presence or absence of any amplification. This strategy would retain the sensitivity benefit of targeting multiple sequences but would lose the ability to distinguish which individual targets amplify. The two IS1081 targets (5' and 3' subregions) already have similar annealing temperatures and could relatively easily be combined into a single assay in this manner. For

all scenarios, LNA-substituted primers may be useful for matching primer melting temperatures across targets (149, 185), given the often limited primer design space for short cfDNA targets.

In addition, while the cost per sample for biotinylated capture probes is low, the up-front cost for new probe designs is high ($200–$400). Design of a modular probe system may be an attractive solution to reduce cost for higher-plex capture. A universal biotinylated adapter sequence could be used to attach non-biotinylated capture sequences. Each capture sequence would contain a region complementary to the universal adapter in addition to a target-specific capture region. The primary limiting factor for the proposed modular design would be co-elution of capture sequences into PCR because they are not immobilized on the bead surface, but it is possible that they could function as a replacement for PCR primers if the 3' carbon spacer that normally blocks extension is removed.

7.4. Ultrashort PCR design

Reducing the PCR amplicon length even further should increase the fraction of short, fragmented cfDNA molecules that contain both primer sites and can be successfully amplified (Figure 7.5). Decreasing the amplicon length has been shown to improve detection sensitivity for other cfDNA targets (4, 78, 79, 81). In urine from a TB-positive patient, decreasing PCR amplicon length from 49 to 39 bp led to a greater than 10-fold increase in detected TB cfDNA concentration, suggesting that even a relatively small decrease in amplicon length may have a disproportionate effect on the detectable fraction of cfDNA. This effect may be particularly pronounced for highly fragmented forms of cfDNA, such as bacterial cfDNA. In our characterization of TB urine cfDNA using NGS (Chapter 6), we found that TB-derived urine cfDNA is significantly shorter than human-derived urine cfDNA. The concentration of TB-derived cfDNA increased exponentially with decreasing fragment length, with a peak fragment length of 19 bp for all TB-positive samples tested. Because concentrations were highest for the shortest fragments, there may still be a substantial sensitivity benefit to be gained by decreasing the amplicon length <40 bp, especially for low concentration samples.

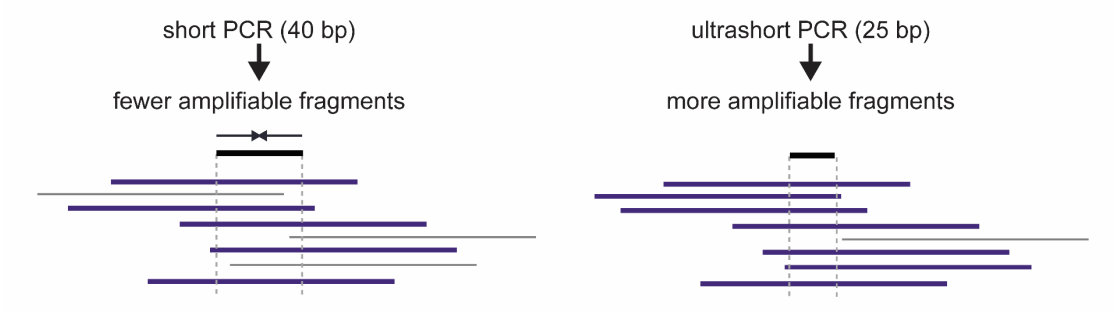


Figure 7.5: Rationale for ultrashort PCR to increase sensitivity. *Detectable fragments are shown in purple.*

7.4.1. *Design of ultrashort PCR assay*

In conventional PCR, the minimum amplicon length is limited to approximately the combined length of the two primers (40 bp in the case of the current assay design). I have designed an ultrashort PCR assay targeting a 25 bp region of IS6110 that uses a stem-loop primer to enable amplification of shorter targets than possible with conventional PCR. The design is based off that described in Shekhtman *et al.* for fetal cfDNA, (79) and mimics a method used for RT-PCR of microRNA (186). The stem-loop primer design and ultrashort PCR amplification mechanism are shown in Figure 7.6. Primer and probe sequences are given in Table 7.2. During the first stage of PCR (10 cycles), a low annealing temperature (45°C) is used to enable partial hybridization of the forward stem-loop primer (10 bp binding region) and reverse primer (12 bp binding region) for pre-amplification of a longer intermediate product. During the second stage of PCR (40 cycles), the annealing temperature is raised (59°C) for conventional PCR amplification of the intermediate product using a universal forward primer, which shares the sequence of the loop sequence of the first stage stem-loop primer, and the full length of the original reverse primer. Amplification is detected using a short sequence-specific minor groove binder (MGB) TaqMan probe (13 bp). The MGB moiety attached to the TaqMan probe enhances affinity so that a shorter probe can be used compared to conventional TaqMan probes. During the second stage of PCR (40 cycles), the stem-loop structure serves to reduce affinity of the first stage primer for the target, avoiding competition for binding with the overlapping MGB TaqMan probe. An iso-dC modified base within the loop of the primer blocks polymerase extension so the intermediate product lacks the stem-loop structure. The relative concentrations of the stem-loop primer (50 nM) and MGB TaqMan probe (100 nM) are important to avoid competition between these sequences during their respective stages in PCR.

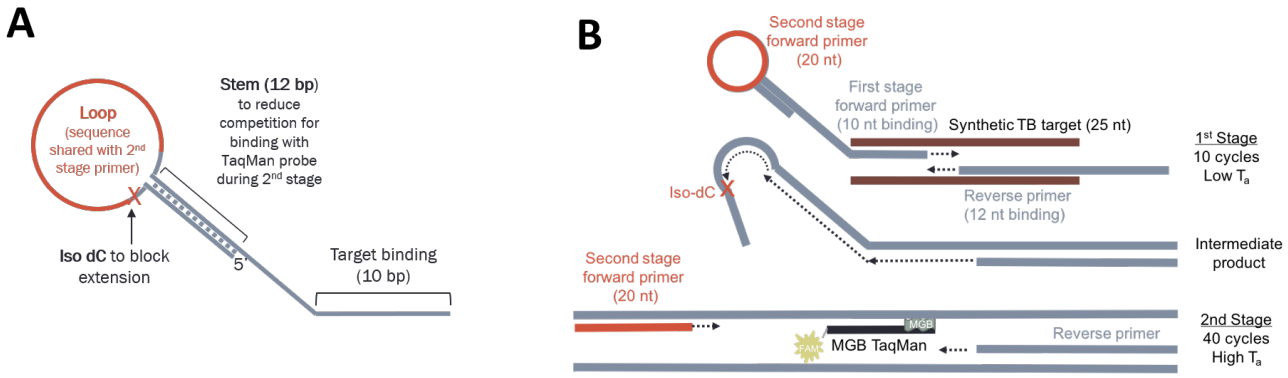


Figure 7.6: Ultrashort PCR design for 25 bp target. (A) *First stage stem-loop primer design.* ***(B)*** *Overview of two-stage PCR amplification.*

Table 7.2: Probe, primer, and target sequences for ultrashort 25 bp IS6110 target.

Oligonucleotide	Sequence
Capture probe for 25 bp target	5’ - /52-Bio/AAAAAAAAAAAAAAAAAAAAGGTGAGGTCTGCTAC/3SpC3/ - 3’
25 bp IS6110 target	5’ – CCGGCTGTGGGTAGCAGACCTCACC – 3’
1st stage stem-loop forward primer	5’ – GCGTAAGAAT/iMe-isodC/AAACGTCGCTCAACTTCCATT CTTACGC**CCGGCTGTGG** – 3’
2nd stage universal forward primer	5’ – AACGTCGCTCAACTTCCATT – 3’
Reverse primer	5’ – TTAGAGAA**GGTGAGGTCTGC** – 3’
MGB TaqMan detection probe	5’ – 6FAM/CCGGCTGTGGGTA/MGBNFQ – 3’

Target/probe binding regions are underlined. First stage ultrashort PCR primer binding regions are bolded. All oligonucleotides were ordered HPLC-purified from IDT, except for the MGB TaqMan probe, which was from Thermo Fisher.

Sequence-specific purification of the 25 bp target was carried out as described in Chapter 3, using 50 pmol of the single capture probe per 50 µL beads. The ultrashort PCR protocol given below:

1. Combine entire neutralized hybridization output (~24 µL) with 1.25 U Hot Start Taq DNA Polymerase (NEB), 1X NEB Standard Taq Buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3) supplemented with an additional 0.5 mM $MgCl_2$ and 70 mM Tris-HCl, 0.8 mM dNTPs (NEB), 50 nM first stage hairpin forward primer, 700 nM second stage universal forward primer, 700 nM reverse primer, and 100 nM MGB TaqMan probe (Table 7.2) in total volume of 50 µL.
2. Amplify in CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with an initial denaturation phase (94°C for 5 min), 10 pre-amplification cycles to extend the first stage loop primer (94°C for 30s and 45°C for 1 min), and 40 amplification cycles (94°C for 30s and 59°C for 1 min).
3. Determine C_q values at a threshold of 100 RFU and calculate recovered copies by a standard curve from 10 – 10^5 copies.

7.4.2. *Preliminary testing of ultrashort PCR*

Sequence-specific purification combined with ultrashort PCR recovered an average of 84% of 25 bp DNA spiked into 10 mL urine (Figure 7.7A). In urine pooled from eight TB-positive individuals, 25 bp PCR detected similar amounts of IS6110 cfDNA compared to 40 bp PCR (Figure 7.7B). Overall, the ultrashort 25 bp PCR assay was less robust than the conventional 40 bp PCR assay, suffering from two limitations: increased susceptibility to PCR inhibition and a lack of specificity (Figure 7.7C-D).

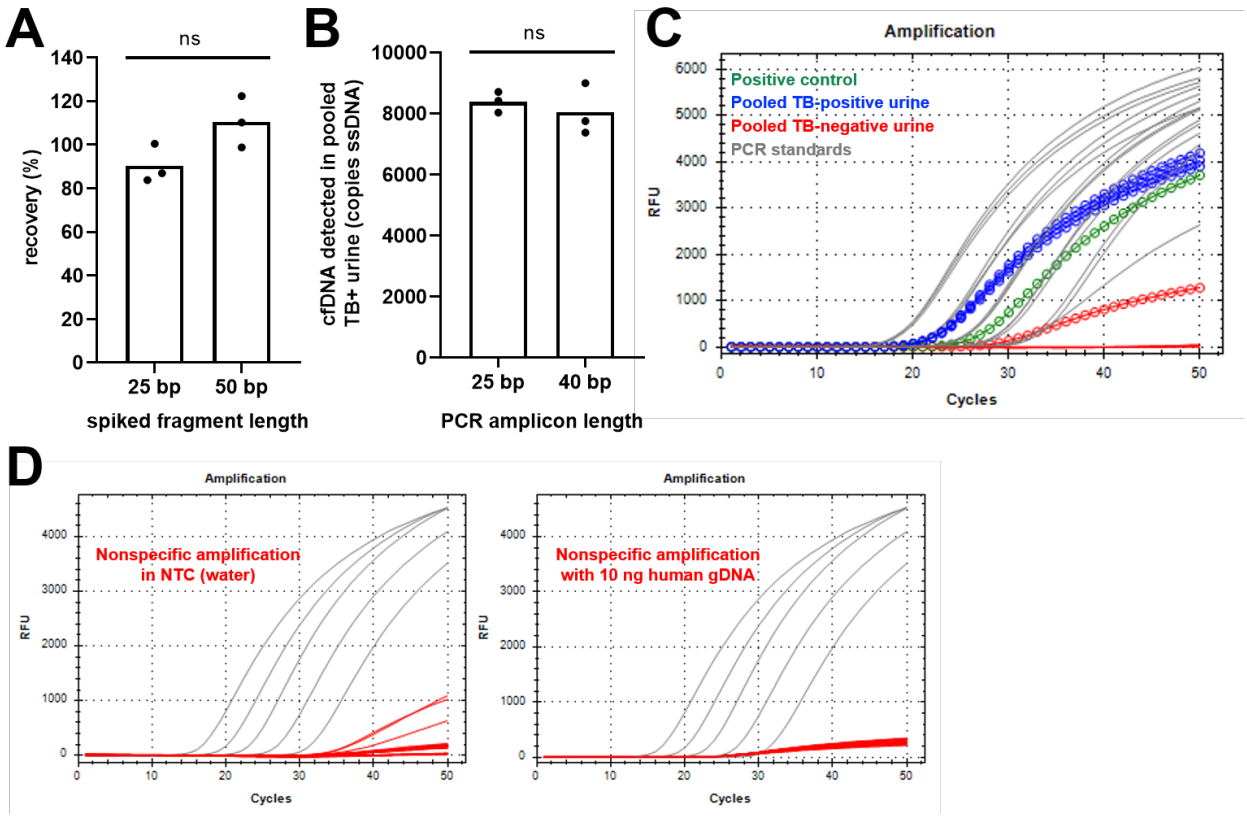


Figure 7.7: Ultrashort PCR performance. (A) *Percent recovery of 1000 copies spiked 25 bp and 50 bp DNA using sequence-specific purification (mean, n=3).* ***(B)*** *Detected copies of IS6110 cfDNA in pooled urine from eight TB-positive individuals (mean, n=3). Note that the 25 bp assay is only capable of recovering a single strand of cfDNA, while the 40 bp assay can recover both strands; a similar detected copy number of ssDNA indicates at least 2-fold more 25 bp cfDNA than 40 bp cfDNA.* ***(C)*** *PCR amplification curves for detection of 25 bp target in pooled TB-positive urine and pooled TB-negative urine. Note two limitations: PCR inhibition (evidenced by reduced slope of cfDNA amplification curves) and nonspecific amplification in TB-negative sample.* ***(D)*** *Nonspecific amplification of water NTC (n=18) and 10 ng human gDNA (n=18).*

Due to the short footprint of the 25 bp target, there is only room for one biotinylated capture probe instead of two as used in the 40 bp assay. Addition of a second probe would result in cross-hybridization of the two probes. As a result, the 25 bp assay recovers only a single strand of cfDNA rather than both

strands. The similar concentrations of 25 bp and 40 bp detected in Figure 7.7B (given as copies of ssDNA) thus indicate at least a 2-fold increase in 25 bp cfDNA compared to 40 bp cfDNA. In the results of this experiment, the increase in 25 bp cfDNA compared to 40 bp cfDNA was enough to compensate for the detection of only a single strand but did not otherwise increase the amount of cfDNA detected.

The first limitation of the ultrashort PCR design is intermittent and variable PCR inhibition in urine cfDNA samples (one representative experiment is shown in Figure 7.7B). It is possible that more 25 bp cfDNA was recovered in this experiment, but the PCR inhibition results in ambiguity for quantification. If the calculated starting quantity of 25 bp cfDNA was corrected for the lower recovery of the positive control seen in this specific experiment (47% for 25 bp vs 80% for 40 bp), the calculated starting quantity of 25 bp cfDNA rises from 8392 to 14401 ssDNA copies. If, instead, regression analysis was used to calculate starting quantity rather than a cycle threshold analysis, as recommended to improve quantification of samples with inhibition (187), the calculated starting quantity of 25 bp cfDNA rises from 8392 to 11470 ssDNA copies. While it is difficult to estimate, based on these analyses I approximate that 25 bp TB cfDNA is present at 2–3.6X the rate of 40 bp TB cfDNA. While developing the ultrashort 25 bp PCR assay, I found that supplementing the Tris-HCl concentration (to 80 mM final concentration, matching that of the 40 bp PCR) in the PCR buffer was necessary for optimal amplification. Further increasing the Tris-HCl concentration to 160 mM reduced but did not eliminate the effects of PCR inhibition in urine cfDNA samples. Varying the pH of urine cfDNA eluate, however, had no effect on inhibition, so it remains unclear what the root cause is.

The second limitation of the ultrashort PCR design is insufficient specificity. Although a sequence-specific TaqMan probe is used for detection, only three nucleotides are unique to the probe and not contained in the primer sequences. To allow binding and extension of short (10–12 bp) sequences during the first 10 cycles of PCR, a low annealing temperature is used. Together, these two design requirements make it difficult to ensure specific amplification. When using a nonspecific DNA binding dye in place of the TaqMan probe, I consistently observed nonspecific amplification equating to 1–10 copies of the intended target, with a T_m of 77–78°C. Using the TaqMan probe helps to distinguish nonspecific amplification from specific amplification, but the probe still occasionally results in signal from nonspecific amplification, likely due to the small number of bases distinguishing the two cases. Eluate from sequence-specific purification of TB-negative urine cfDNA often leads to nonspecific amplification early enough (equivalent of 10–50 copies of the intended target, Figure 7.7C) that it would interfere with detection of low to moderate concentration TB-positive samples. In PCR NTCs and PCR

spiked with 10 ng human gDNA (Figure 7.7D), the specificity is slightly better. Nonspecific amplification occurs later and/or can be distinguished from specific amplification by the shape of the amplification curve. I attempted to improve specificity by optimizing the annealing temperature during the first 10 cycles of PCR. Surprisingly, PCR efficiency remained high (>90%) well above what would be expected based on the estimated T_ms of the first stage primer binding regions. The estimated T_ms for the first stage binding regions for the forward and reverse primer, respectively, are 43.5°C and 51.8°C (calculated using OligoAnalyzer, which does not incorporate the effect of ssDNA overhangs). The measured T_ms were approximately 55°C and 58°C, respectively. The maximum PCR annealing temperature was 57°C, nearing that used in the second stage of PCR. For PCR NTCs, increasing the first stage annealing temperature did not have an obvious impact on specificity, but I did not test the effect on specificity in eluate from sequence-specific purification of TB-negative urine cfDNA.

7.4.3. *Conclusions and next steps towards ultrashort PCR*

Decreasing the amplicon length from 40 bp to 25 bp did not have as much of an effect on detected concentration as hypothesized, especially considering the NGS fragment length profile of TB cfDNA discussed in Chapter 6. However, the exponential rise of short fragments did not start until <20 bp for most samples. The difference between fragments of ~40 bp and ~25 bp was more subtle (and most fragments will have to be longer than the exact amplicon length to be amplifiable). Given the other challenges of the ultrashort 25 bp PCR assay (lack of specificity, PCR inhibition), it may not be worth continuing to pursue this approach. It is possible that the minor increase in detectable cfDNA for 25 bp PCR compared to 40 bp PCR (estimated 2–3.6X increase) could still benefit the detection of low concentration samples near the limit of detection of PCR. For this to be feasible, specificity would need to be improved, PCR inhibition would need to be resolved, and a design enabling addition of a second capture probe to detect both strands of cfDNA would be required. Increasing first stage annealing temperature may improve specificity in urine cfDNA, even if it had no impact on PCR NTCs. Substitution of LNA bases may be a viable strategy to increase the maximum annealing temperature and/or to reduce the length of primer binding regions (to accommodate more unique bases in the TaqMan probe) or biotinylated capture probes (to accommodate a second capture probe without substantial overlap).

7.5. INCREASED SAMPLE VOLUME

Increasing the input urine sample volume above 10 mL would proportionally increase the copy number of cfDNA targets available for amplification. Unlike the previous two strategies (multiplexing and ultrashort PCR), this approach does not rely on expanding the detectable fraction of cfDNA molecules,

but rather increasing the concentration factor for the current target. For low concentration samples currently near the limit of detection of PCR, scaling up the volume 2–4X may be sufficient to enable more confident detection. A disadvantage of this strategy is that it will not help scale down the volume to ease future assay simplification.

To avoid significantly scaling up assay cost, I attempted to increase the sample volume while keeping the amount of magnetic beads used the same. I scaled all other reagents (NaCl, Tween-20) proportionally to sample volume. To accommodate increased sample volume, I carried out denaturation and hybridization in 50 mL tubes rather than 15 mL tubes. To achieve sufficient denaturation of larger sample volumes, a boiling water bath was required rather than a 95°C water bath (Figure 7.8). Recovery of ssDNA remained high regardless of denaturation protocol, confirming that denaturation was the limiting step. I did not test heat block denaturation because we did not have a heat block insert for 50 mL tubes available.

Figure 7.8B shows the percent recovery for 10, 20, 30, and 40 mL samples using the modified denaturation protocol. Recovery for a 20 mL sample (82%) was similar to or slightly lower than that for a 10 mL sample (100%). Recovery for 30 mL and 40 mL samples decreased moderately (60% and 51%, respectively). This experiment included only a single replicate for each volume, so it is unclear if the difference in recovery for 10 mL and 20 mL samples is significant, but it is within the normal range of assay variation. In an experiment using a less effective denaturation protocol with replicates (Figure 7.8A), recovery was the same for 10 mL and 20 mL samples. Increasing the hybridization time from 30 minutes to 60 minutes did not improve recovery for 40 mL samples (not shown).

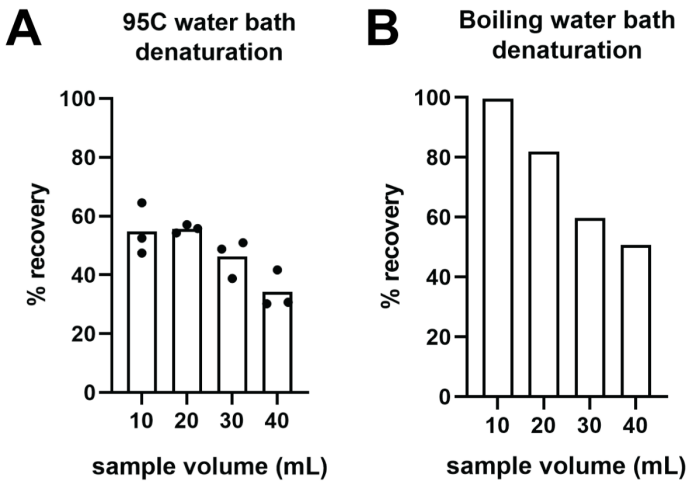


Figure 7.8: Effect of increasing sample volume on percent recovery. *The volume of magnetic beads was kept the same (50 µL) regardless of sample volume.* ***(A)*** *Denaturation in 95°C water bath for 15 minutes results in insufficient denaturation in 50 mL tubes, regardless of sample volume (n=3; 10^3 copies 50 bp*

dsDNA input). ***(B)*** *Denaturation in boiling water bath for 15 minutes effectively denatures larger samples (n=1; 10^3 copies 50 bp dsDNA input).*

Even given the tradeoff of a drop in percent recovery, analysis of larger sample volumes may be worthwhile, as the quantity of cfDNA molecules recovered matters more than the percent recovery. However, is unclear whether an increase in sample volume may lead to worsening of PCR inhibition or bead behavior problems observed in clinical samples from hospital inpatients but not healthy controls. Anecdotally, I observed minor PCR inhibition more frequently with 30 and 40 mL healthy control urine samples than with 10 and 20 mL samples. Based on these preliminary results, I recommend exploring the use of 20 mL samples in future clinical testing. It should also be investigated whether replicate testing versus pooling into larger samples results in more reliable detection of low concentration samples when total sample volume is limited.

7.6. TROUBLESHOOTING BEAD BEHAVIOR

Processing clinical samples from hospital inpatients (many living with HIV and/or severely ill) presented additional challenges not reflected during assay development using healthy control samples. Specifically, some samples led to problems with the behavior of the magnetic beads after the denaturation step, causing bead loss during wash steps and/or bead transfer to PCR during the elution step due to poor accumulation of beads on the magnet. This did not prevent processing of any samples, but occasionally made processing the samples more difficult, would complicate hand-off of the assay to less experienced users, and could potentially cause false negatives due to bead loss, PCR inhibition from beads, or elution of incomplete volume to avoid bead transfer. In cases where wash or elution steps were particularly difficult, I used a minicentrifuge to help pellet the beads before removing the wash buffer or eluate and sometimes left larger residual volumes to avoid bead loss or transfer to PCR. TB cfDNA was still successfully detected in several samples with bead behavior problems, but resolving this issue will improve assay usability and potentially improve sensitivity in samples with proteinuria and/or blood.

7.6.1. *Characterization of sample variability and bead behavior problems*

Visual examples of sample variability and bead behavior problems are shown in Figure 7.9A-E. In addition, there was visible coagulation of some samples after heating to denature cfDNA. Quantification of the incidence of bead behavior problems indicates that problems were more frequent in samples with high levels of protein and/or blood (Figure 7.9F-G). Odds ratios for bead behavior problems in samples with high vs. low/no protein and high vs. low/no blood are shown in Table 7.3. The two categories of bead problems (bead loss during wash steps and bead transfer during elution step) did not

always occur concurrently, and high levels of protein or blood did not guarantee bead problems. There was also occasionally inconsistency in bead behavior across replicates of the same sample, but I only measured protein and blood concentrations for one replicate per sample. In spike-in experiments, as little as 1 µg (0.1 µL) of magnetic beads led to noticeable PCR inhibition.

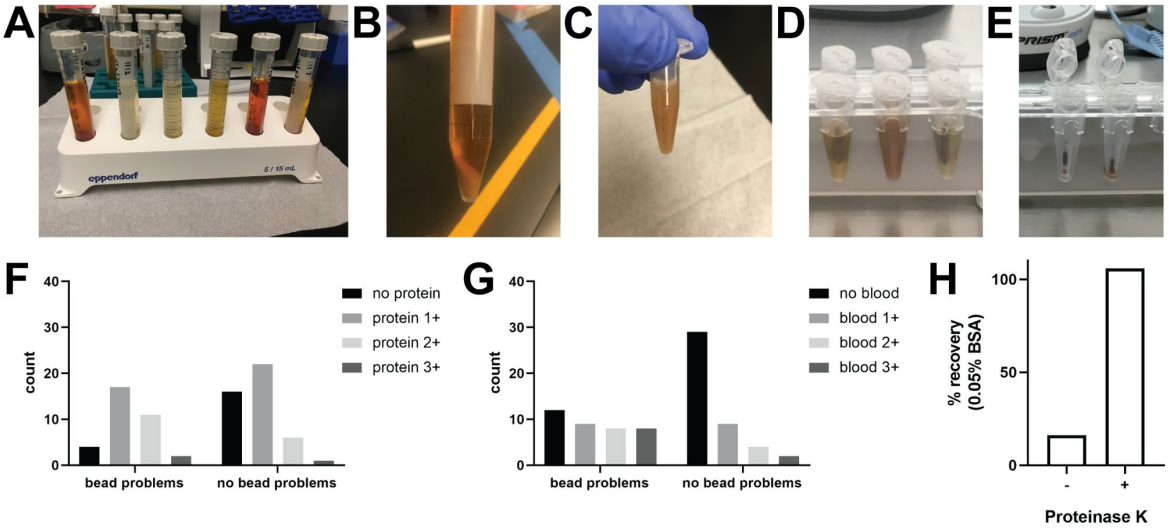


Figure 7.9: Characterization of sample variability and bead processing problems. *Visual examples of sample variability, including color* ***(A)*** *and amount of sediment* ***(B)****, and examples of bead behavior problems, including coagulation after heating* ***(C)****, poor aggregation on the magnet during wash steps* ***(D)****, and poor aggregation on the magnet during elution* ***(E).*** *Incidence of bead behavior problems (bead loss during wash steps and/or bead transfer during elution) based on the amount of protein* ***(F)*** *and blood* ***(G)*** *measured in urine samples from hospital inpatients in South Africa using Fisherbrand 10SG Urine Reagent Strips (healthy controls in USA excluded).* ***(H)*** *Addition of Proteinase K digestion step eliminates bead behavior problems and increases recovery in the presence of 0.05% crude BSA.*

Table 7.3: Odds ratios for bead behavior problems based on presence of proteinuria or blood.

		Odds ratio (95% CI)	P-value
Bead loss during wash steps	**Protein**	4.0 (1.2 – 11.3)	0.030
	Blood	3.5 (1.0 – 11.5)	0.071
Bead transfer during elution	**Protein**	2.6 (0.83 – 10.5)	0.15
	Blood	3.2 (0.89 – 10.5)	0.097
Any bead problems	**Protein**	3.3 (1.1 – 10.8)	0.050
	Blood	3.6 (1.0 – 11.6)	0.059

Odds ratios calculated for protein measurement of ≥100 mg/dL vs <100 mg/dL and blood measurement of ≥"++" vs <"++" using Fisherbrand 10 SG Urine Reagent Strips. The Baptista-Pike method was used to calculate 95% Cis and Fisher's exact test was used to calculate p values.

7.6.2. *Addition of proteinase K digestion step*

Although the full complexity of what variables contribute to bead behavior problems remains unclear, it appears that proteinuria plays a role. This is not an unexpected finding, as adsorption of denatured

proteins to the bead surface after heating may cause bead aggregation. I found that spiking in crude BSA (but not pure molecular biology grade BSA) somewhat replicated the problem. Up to 0.3% crude BSA, there were no difficulties processing samples. At 0.5% crude BSA, it was somewhat more difficult to resuspend beads after centrifugation prior to wash steps due to coagulation of the solution. There was no difficulty during the wash steps, but significant difficulty during the elution step. Beads had a tendency to transfer to PCR, and the full elution volume could not be transferred. The calculated percent recovery was 16% (Figure 7.9H). For >0.5% crude BSA, wash steps could not be completed due to lack of bead aggregation on the magnet.

I next tested whether the addition of a proteinase K digestion step prior to denaturation and hybridization could avoid bead behavior problems in samples with high protein concentrations. I added 120 µL of 20 mg/mL proteinase K (NEB) to the urine sample simultaneously with the hybridization reagents (beads, NaCl, Tween-20) and incubated at 65°C for 20 minutes. Deactivation of proteinase K was then combined with the existing cfDNA denaturation step at 95°C for 15 minutes. Surprisingly, the streptavidin-coated magnetic beads with immobilized biotinylated probes were resistant to proteinase K digestion and could be added prior to the denaturation step without a loss in performance. Using this digestion protocol eliminated the bead handling problems described above in the presence of 0.5% crude BSA, leading to a calculated percent recovery of 106%, over 6-fold more than without proteinase K (Figure 7.9H). I did not optimize the proteinase K concentration, digestion temperature, or digestion time, and did not test effectiveness of the digestion step with higher BSA concentrations.

In a single clinical urine sample, addition of a proteinase K digestion step eliminated problems with bead transfer to PCR during elution and reduced PCR inhibition but did not affect the detected TB cfDNA concentration. Proteinase K digestion also eliminated urine sediment, which was otherwise moderate for this sample, without the need for an initial centrifugation step to isolate the cell-free urine fraction. Proteinase K digestion may have the potential to replace the initial centrifugation step used to remove urine sediment that would otherwise be concentrated alongside beads after hybridization. Additional volume of known problematic samples was unavailable, so I was unable to complete a thorough evaluation of the effects of proteinase K digestion on the ease of sample processing and detection sensitivity in clinical samples.

7.6.3. *Proteinase K digestion protocol*

Changes from original denaturation and hybridization protocol are bolded.

1. Add **120 µL 20 mg/mL proteinase K**, 2.5 mL 5 M NaCl, 127 µL 10% Tween-20, and 50 µL prepared beads to each 10 mL urine sample. If spiking in positive control, add it now.
2. Mix well by inversion.
3. **Digest for 20 minutes in dry bath with 15 mL tube block preheated to 65°C.**
4. Denature cfDNA and **deactivate proteinase K** for 15 minutes in a dry bath with 15 mL tube block preheated to 120°C. Alternatively, a 95°C water bath can be used.
5. Rotate for 30 minutes at room temperature to hybridize target cfDNA to capture probes.

7.6.4. *Conclusions and next steps*

Addition of a proteinase K digestion step appears likely to mitigate the effects of sample variability, with minimal impact on the overall assay workflow. In addition to improving the ease and consistency of sample processing and making the assay more amenable to adoption by new users, this change may increase sensitivity by enabling more reliable detection in samples that would otherwise suffer from bead loss or PCR inhibition. Optimization of the digestion step may allow for a reduced incubation time and/or cost (by either reducing proteinase K volume or using less costly proteinase K). If proteinase K digestion can replace the initial centrifugation step, which is intended to isolate the cell-free urine fraction and avoid sediment that can interfere with downstream processing, it may help decrease manual steps, processing time, and sample loss during liquid transfer steps. Although there does not seem to be any downside to including a digestion step, testing is needed to gauge its effectiveness across varied clinical urine samples. I recommend considering implementation of proteinase K digestion in future clinical testing.

7.7. CONCLUSIONS

Overall, multiplexed capture was the most promising strategy for increasing assay sensitivity, although there may be incremental benefits to be gained from ultrashort PCR and increased sample volume. Specifically, IS6110 and IS1081 should be prioritized for inclusion in future development of a multiplexed assay. Addition of a proteinase K step needs more validation but has strong potential to lessen the effects of sample variability, improve assay usability, and possibly increase sensitivity.

8. CONCLUSIONS AND FUTURE DIRECTIONS

The overall goal of my thesis project was to improve the diagnosis of TB from urine cfDNA. We hypothesized that we could increase the sensitivity of TB diagnosis from urine cfDNA by improving the

retention of short, dilute cfDNA fragments during the sample preparation step using sequence-specific purification. To develop a method capable of detecting TB-specific cfDNA in clinical urine samples with high sensitivity, I had to invest significant effort in optimizing its analytical performance. The resulting assay has <1 copy/mL sensitivity from 10 mL urine samples and the highest diagnostic accuracy for a TB urine cfDNA assay reported to date. The results of the studies described here demonstrate the promise of our sequence-specific purification approach and will lay the foundation for future clinical validation of our assay. In addition, my dissertation work provides several important contributions to the growing evidence for urine cfDNA as a diagnostic target for TB, in the forms of both clinical evidence and molecular characterization, and collectively motivates and informs continued development of urine cfDNA-based assays. A sensitive urine cfDNA-based test built on sequence-specific purification could significantly contribute to improving sample availability and expanding access to rapid TB diagnostics.

While my research primarily focused on TB cfDNA, another overarching theme throughout my graduate work was the desire to make my research, and the methods developed from it, relevant, useful, and readily accessible to as wide of an audience as possible. To achieve this, I published a detailed, user-ready protocol for the TB urine cfDNA assay and wrote several of my published papers to target a broad audience interested in cfDNA, including applications beyond TB. I hope that the methods described here can continue to be used to investigate the myriad applications of cfDNA.

Key results and conclusions from each segment of my dissertation work are outlined below.

8.1. Development of sequence-specific purification method for urine cell-free DNA

With the goal of developing an improved diagnostic assay for TB urine cfDNA, I designed and optimized a sensitive hybridization method that uses sequence-specific oligonucleotide capture probes immobilized on magnetic beads to improve extraction of short cfDNA from large-volume urine samples. By combining sequence-specific purification with qPCR targeting a short (40 bp) region of the MTB-specific insertion sequence IS6110, we can detect TB-derived cfDNA in urine.

The resulting protocol achieves the following performance metrics:

- Near 100% recovery of short (25–150 bp) DNA spiked into 10 mL urine, independent of fragment length and concentration
- Limit of detection of ≤5 copies of dsDNA in 10 mL urine (0.5 copies/mL)
- Amplification of cfDNA from 10 mL urine in a single PCR well (500X concentration)

Several key innovations contribute to the strong performance of this method:

- Dual biotinylated capture probes ensure consistent, high recovery by facilitating optimal probe density on the bead surface, improving thermostability of the probe-bead linkage, and eliminating interference by endogenous biotin.
- A two-probe system enables recovery of both strands of double-stranded DNA.
- Surface-based hybridization accommodates larger sample volume than solution-based hybridization, while remaining cost-effective.
- LNA-substituted primers allow for precise tuning of primer melting temperatures, despite limited primer design space. Increasing the PCR annealing temperature to slightly above the primer melting temperature encourages specific amplification.

8.2. Analytical comparison of urine cell-free DNA extraction methods

To build a better understanding of existing methods for urine cfDNA extraction, I compared the analytical performance of several lab-based protocols and commercially available kits. This work was intended to go beyond the scope of TB, appeal to broader clinical audiences interested in urine cfDNA, and help inform selection of optimal urine cfDNA extraction methods. Urine cfDNA extraction methods varied widely in ability to capture short, dilute urine cfDNA. Employing sub-optimal methods may compromise clinical results due to low recovery, fragment length bias, dependence on urine composition, or PCR inhibition.

Below are the conclusions for each urine cfDNA extraction method included in the comparison:

- Our in-house sequence-specific purification method outperforms alternate urine cfDNA extraction methods by improving recovery of short fragments, motivating continued development of our hybridization-based TB urine cfDNA assay.
- The Wizard/GuSCN method, which was used in previous TB urine cfDNA studies, has limited overall recovery (<17%) and is highly dependent on urine pH and background DNA concentration. It is likely to fail in a subset of clinical urine samples with non-optimal conditions.
- The Q Sepharose method improves upon the performance of the Wizard/GuSCN method, with high recovery (63–75%) of fragments down to 40 bp. Recovery was still reduced (<10%) for the shortest fragments (25 bp). This method is an ideal candidate for extraction of total urine cfDNA prior to sequencing, which is not possible using sequence-specific purification.

- The Norgen Urine Cell-Free Circulating DNA Purification Kit had high recovery (72%) of the shortest (25 bp) fragments but moderate recovery (30–41%) of longer fragments. It led to consistent PCR inhibition that makes quantification unreliable.
- The Qiagen QIAamp Circulating Nucleic Acid Kit did not perform as well in urine as has been previously reported for plasma. It had low (18%) recovery of the 150 bp fragment, with almost no recovery of shorter fragments (25–80 bp).
- The Thermo Fisher MagMAX Cell-Free DNA Isolation Kit had recovery inversely proportional to fragment length. It had high recovery (66%) of the 150 bp fragment but almost no recovery of the 40 bp fragment.

8.3. Detection of tuberculosis cell-free DNA in clinical urine samples

Using sputum Xpert MTB/RIF as the reference standard, our TB urine cfDNA assay had 83.7% sensitivity (95% CI: 71.0–91.5%) and 100% specificity (95% CI: 86.2–100%) for diagnosis of active pulmonary TB in adults in a clinical cohort study in South Africa (n=74). This is the highest reported diagnostic accuracy for a TB urine cfDNA test to date. These results demonstrate that urine cfDNA is a viable biomarker for TB diagnosis and support our hypothesis that increasing the recovery of short cfDNA fragments during sample preparation improved the sensitivity of TB urine cfDNA detection.

Additional important results from clinical testing of our TB urine cfDNA assay are as follows:

- The detected IS6110 cfDNA concentration was low and variable (range 0.14–2804 copies/mL; median 14.6 copies/mL) and was inversely correlated with CD4 count and days to culture positivity.
- Sensitivity was non-significantly higher in HIV-positive (88.2%) compared to HIV-negative patients (73.3%) but was not dependent on CD4 count. The potential to diagnose patients regardless of HIV status and CD4 count is a critical advantage of urine cfDNA over urine LAM.
- Sensitivity remained high in sputum smear-negative (76.0%) and urine LAM-negative (76.5%) patients. Urine cfDNA has the potential to diagnose patients that may be missed by these alternate diagnostic methods.

8.4. Next-generation sequencing of tuberculosis urine cell-free DNA

To better characterize TB-derived cfDNA in urine, validate its use as a biomarker for TB, and inform the development of improved TB urine cfDNA diagnostic assays, we investigated the size and composition of

TB-derived urine cfDNA using NGS (n=12). To reduce bias against highly fragmented urine cfDNA, we used a combination of DNA extraction (Q sepharose) and single-stranded sequence library preparation methods (SRSLY) demonstrated to recover short, degraded cfDNA fragments.

The key findings of our sequencing analysis are as follows:

- TB-derived urine cfDNA is significantly more degraded than human-derived urine cfDNA, with median fragment lengths of ≤19–45 bp and 42–91 bp, respectively.
- TB-derived cfDNA increases in abundance exponentially with decreasing fragment length, with a peak fragment length of ≤19 bp in all samples (the minimum detectable fragment length using our analysis methods). This result motivates the continued development of methods (sample preparation and amplification) capable of targeting shorter fragments to improve the sensitivity of TB urine cfDNA assays.
- Urine cfDNA fragments span the TB genome with relative uniformity. No *de novo* diagnostic targets over-represented in urine cfDNA were identified. Targeting known multicopy genomic elements remains the most promising strategy for inherent signal amplification.
- Our methodology revealed a larger fraction of short human genomic cfDNA than previously reported, with peak fragment lengths of 29–53 bp. The combination of Q Sepharose DNA extraction and single-stranded library preparation may prove useful for other applications requiring analysis of highly fragmented forms of DNA.

8.5. Multiplexing and next steps towards continued assay development

Despite the strong analytical and clinical performance of the TB urine cfDNA assay, the concentrations of TB cfDNA detected in clinical urine samples were low. There remains opportunity to increase assay sensitivity and robustness through more reliable detection of low concentration samples. In the final portion of my dissertation work, I explored several strategies for continued assay improvement, with the objective of determining which strategies were most promising for future development:

- Multiplexing was identified as the most promising strategy for increasing assay sensitivity. Multiplexed capture of 2–5 targets was feasible without assay modifications. Preliminary designs targeting two different subregions of multicopy element IS1081 performed well and should be prioritized for combination with IS6110 in a multiplexed assay.
- Ultrashort PCR (25 bp) increased the amplifiable fraction of TB cfDNA molecules by only a moderate amount (estimated 2–4X) compared to 40 bp PCR. The 25 bp assay also suffered from

several limitations (capture of only a single strand of cfDNA, lack of specificity, and susceptibility to PCR inhibition) that would need to be resolved if it is to be implemented.

- Increasing the input sample volume from 10 mL to 20 mL had minimal effect on percent recovery (80–100%) and would double the quantity of TB cfDNA molecules available for detection. Recovery remained moderately high (50–60%) for even larger samples (30–40 mL).
- Addition of a proteinase K digestion step has the potential to mitigate the effects of sample variability (proteinuria, blood) on bead handling, with minimal impact on assay workflow, and should be considered for implementation in future clinical testing. It may also be able to replace the initial centrifugation step required to remove urine sediment.

8.6. KEY ISSUES TO BE ADDRESSED IN FUTURE WORK

8.6.1. *Expanded clinical testing*

Additional clinical studies with larger sample sizes and expanded scope are needed to better validate our TB urine cfDNA and more conclusively compare sensitivity across subgroups. In particular, future studies should include Xpert-negative, culture-positive participants and expand enrollment to include outpatients. Studies focusing on patient populations who are underserved by existing rapid diagnostics, including children, people living with HIV, and individuals with EPTB, will be critical for establishing the potential impact of urine cfDNA testing. Specificity should be more thoroughly tested through both inclusion of larger control groups and targeted analytical testing with spiked nontuberculous mycobacteria (NTM). Finally, our assay could be also used to investigate important questions remaining about TB urine cfDNA, including its optimal storage conditions, potential for treatment monitoring, and correlation with bacterial load.

8.6.2. *Development of multiplexed assay*

I have developed several new assays (capture probes and PCR) targeting multicopy elements within the TB genome, including those present in MTB strains lacking IS6110. Future work should focus on improving specificity of the new PCRs and combining targets into a multiplexed assay. While multiplexed capture should be relatively straightforward, multiplexed amplification of short targets without the footprint for sequence-specific detection probes may be more challenging. Some strategies for short-target multiplexed PCR are outlined in Section 7.3.4.

8.6.3. *Comparison of TB cfDNA in urine and blood*

My work focused specifically on the study of TB cfDNA in urine, but TB cfDNA is also present in blood (61, 62, 164). We initially selected urine due to the ease of collection and large available sample volume. For some patients, especially children, collection of a blood sample may be more straightforward than collection of a urine sample. The concentrations of total cfDNA in urine and blood are expected to be similar, but concentrations and fragment lengths of TB-derived cfDNA have not been directly compared across the two sample types. It is possible that TB cfDNA is more intact in blood than in urine, which may make it easier to detect. Future work should characterize the fragment length distribution of circulating TB cfDNA in blood using NGS, adapt the sequence-specific purification method for whole blood, plasma, and/or serum samples, and compare detection sensitivity of our cfDNA assay in matched urine and blood samples. If successful, detection of cfDNA in blood should be considered for prioritization in future clinical testing for pediatric TB.

8.6.4. *Simplification to format more suitable for use in resource-limited settings*

The goal of my thesis project was to design a maximum sensitivity lab-based assay for TB urine cfDNA, which currently requires a trained user, laboratory equipment, and significant hands-on time. For the urine cfDNA assay to compete with existing rapid TB tests, substantial simplification of the assay format will be necessary. In the future, our lab-based test should be simplified into a format more suitable for use in resource-limited settings, possibly by adapting technologies in development for silica-based magnetic bead purification for sequence-specific purification. Promising technologies for assay simplification include those that employ immiscible phase filtration (188–191) and high-gradient magnetic separation (192). Because existing technologies cannot accommodate large input sample volumes, the 10 mL volume required for our assay may pose a significant challenge for assay simplification. Ideally, further improvements in the ability to detect low concentration samples (for example, through multiplexing) will allow for a reduction in sample volume. Alternatively, magnetic beads can be concentrated manually by centrifugation or a semi-automated device (such as the nRichDx system (193)) prior to processing on a downstream automated device.

8.7. List of publications & presentations

8.7.1. *Publications included in this dissertation*

- **Oreskovic A**, Brault ND, Panpradist N, Lai JJ, Lutz BR. Analytical comparison of methods for extraction of short cell-free DNA from urine. Journal of Molecular Diagnostics. 21(6):1067-78 (2019). https://doi.org/10.1016/j.jmoldx.2019.07.002.
- **Oreskovic A**, Lutz BR. Ultrasensitive hybridization capture: reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. PLOS ONE 16(2): e0247851 (2021). https://doi.org/10.1371/journal.pone.0247851.
- **Oreskovic A**, Panpradist N, Marangu D, Ngwane MW, Magcaba ZP, Ngcobo S, Ngcobo Z, Horne DJ, Shapiro AE, Wilson DPK, Drain PK, Lutz BR. Diagnosing pulmonary tuberculosis using sequence-specific purification of urine cell-free DNA. Journal of Clinical Microbiology 59(8):e00074-21 (2021). https://doi.org/10.1128/jcm.00074-21.
- **Oreskovic A**, Waalkes A, Holmes EA, Rosenthal CA, Wilson DPK, Shapiro AE, Drain PK, Lutz BR, Salipante SJ. Unbiased sequencing of *Mycobacterium tuberculosis* urinary cell-free DNA reveals extremely short fragment lengths (manuscript in preparation for submission to Journal of Clinical Microbiology).

8.7.2. *Other publications*

- Panpradist N, Wang Q, Ruth P, Kotnik K, **Oreskovic A**, Miller A, Stewart S, Vrana J, Han P, Beck I, Starita L, Frenkel L, Lutz B. Simpler and faster Covid-19 testing: strategies to streamline SARS-CoV-2 molecular assays. EBioMedicine 64: 103236 (2021). https://doi.org/10.1016/j.ebiom.2021.103236.
- Panpradist N, Atkinson RG, Roller M, Kline E, Wang Q, Hull IT, Kotnik JH, **Oreskovic A**, Bennett C, Leon D, Lyon V, Han PD, Starita LM, Thompson MJ, Lutz BR. Harmony COVID-19: a ready-to-use kit, low-cost detector, and smartphone app for point-of-care SARS-CoV-2 RNA detection (manuscript under review at Science Advances).

8.7.3. *Protocols*

- Ultrasensitive hybridization capture of short tuberculosis cell-free DNA from urine. Oreskovic A, Lutz B. http://dx.doi.org/10.17504/protocols.io.bep4jdqw.

8.7.4. *Conference presentations*

- Sequence-specific purification of urine cell-free DNA for tuberculosis diagnosis. Biomedical Engineering Society Annual Meeting. Phoenix, AZ. October 13, 2017. Oral presentation.
- Sensitive hybridization capture and detection of urine cell-free DNA for tuberculosis diagnosis. Association for Molecular Pathology Annual Meeting. Baltimore, MD. November 9, 2019. Poster presentation.
- Sequence-specific hybridization capture of urine cell-free DNA to diagnose pulmonary tuberculosis. The 51st Union World Conference on Lung Health. Virtual. October 21, 2020. Oral presentation.

9. APPENDIX A: PROTOCOLS AND OLIGONUCLEOTIDE SEQUENCES

9.1. OLIGONUCLEOTIDE SEQUENCES

Table 9.1: List of oligonucleotide sequences used for detection of TB cfDNA.

40 bp IS6110 target	
Oligo	**Sequence**
IS6110 target (40 bp)	5'-CGAACCCTGCCCAGGTCGACACATAGGTGAGGTCTGCTAC-3'
Reverse complement of IS6110 target (40 bp)	5'-GTAGCAGACCTCACCTATGTGTCGACCTGGGCAGGGTTCG-3'
Synthetic positive control (50 bp)	5'-CGAACCCTGCCCAGGTCGACA**CCATTCAACA**CATAGGTGAGGTCTGCTAC-3'
Reverse complement of positive control (50 bp)	5'-GTAGCAGACCTCACCTATG**TGTTGAATGG**TGTCGACCTGGGCAGGGTTCG-3'
Biotinylated probe #1 (BP1, targets positive control)	5'-/52-Bio/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/-3'
Biotinylated probe #2 (BP2, targets reverse complement)	5'-/52-Bio/AAAAAAAAAAAAAAAAAAAACCCTGCCCAGGTCGA/3SpC3/-3'
Forward primer	5'-CGAACCCTGCCCAGGTCGA-3'
LNA-substituted reverse primer (recommended)*	5'- GTA+GCAGA+CCTCACCTATGTGT-3'
Reverse primer (no LNA; not recommended)*	5'- GTAGCAGACCTCACCTATGTGT-3'
80 bp target (length dependence experiments)	5'-CGAACCCTGCCCAGGTCGACA**CCATTCAACA**CATAGGTGAGGTCTGCTACACACCATTCAATTTCATCACTGCCAATACT-3'
150 bp target (length dependence experiments)	5'-CGAACCCTGCCCAGGTCGACA**CCATTCAACA**CATAGGTGAGGTCTGCTACACACCATTCAATTTCATCACT

	GCCAATACTCCACTCTCATCTACACAACCCATTAGTACCTTACCTCGCTTCCTATCCCAATTCACTTAATCTT AAACCG-3’
25 bp IS6110 target	
Oligo	**Sequence**
IS6110 target and positive control (25 bp)	5’-CCGGCTGTGGGTAGCAGACCTCACC-3’
Biotinylated probe	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAAGGTGAGGTCTGCTAC/3SpC3/-3’
1st stage stem-loop forward primer	5’ – GCGTAAGAAT/iMe-isodC/AAACGTCGCTCAACTTCCATT CTTACGCCCGGCTGTGG – 3’
2nd stage forward primer	5’ – AACGTCGCTCAACTTCCATT – 3’
Reverse primer	5’ – TTAGAGAAGGTGAGGTCTGC – 3’
MGB TaqMan probe	5’ – 6FAM/CCGGCTGTGGGTA/MGBNFQ – 3’
IS1081 target (5’ subregion)	
Oligo	**Sequence**
IS1081 5’ target (41 bp)	5’-TCGCGTGATCCTTCGAAACGCGACCAGCAAAAGTCAATCGA-3’
Reverse complement of IS1081 5’ target (41 bp)	5’-TCGATTGACTTTTGCTGGTCGCGTTTCGAAGGATCACGCGA-3’
Synthetic positive control (51 bp)	5’-TCGCGTGATCCTTCGAAACG**ACACCATTCA**CGACCAGCAAAAGTCAATCGA-3’
Reverse complement of positive control (51 bp)	5’- TCGATTGACTTTTGCTGGTCG**TGAATGGTGT**CGTTTCGAAGGATCACGCGA-3’
Biotinylated probe #1 (BP1, targets positive control)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAAATTGACTTTTGCTGGTCG/3SpC3-3’
Biotinylated probe #2 (BP2, targets reverse complement)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAACGTGATCCTTCGAAACG/3SpC3-3’
Forward primer	5’-TCGCGTGATCCTTCGAAACG-3’
Reverse primer	5’-TCGATTGACTTTTGCTGGTCG-3’
IS1081 target (3’ subregion)	
Oligo	**Sequence**
IS1081 3’ target (40 bp)	5’-CCGAATCAGTTGTTGCCCAATATGATCGGGTACTCGACGC-3’
Reverse complement of IS1081 3’ target (40 bp)	5’- GCGTCGAGTACCCGATCATATTGGGCAACAACTGATTCGG-3’
Synthetic positive control (50 bp)	5’-CCGAATCAGTTGTTGCCCA**ACACCATTCA**ATATGATCGGGTACTCGACGC-3’
Reverse complement of positive control (50 bp)	5’- GCGTCGAGTACCCGATCATAT**TGAATGGTGT**TGGGCAACAACTGATTCGG-3’
Biotinylated probe #1 (BP1, targets positive control)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAATCGAGTACCCGATCATA/3SpC3-3’

Biotinylated probe #2 (BP2, targets reverse complement)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAAAATCAGTTGTTGCCCA/3SpC3-3’
Forward primer	5’-CCGAATCAGTTGTTGCCCA-3’
Reverse primer	5’-GCGTCGAGTACCCGATCATA-3’
Direct repeat (DR) target	
Oligo	**Sequence**
DR target (36 bp)	5’-GTTTCCGTCCCCTCTCGGGGTTTTGGGTCTGACGAC-3’
Reverse complement of DR target (36 bp)	5’- GTCGTCAGACCCAAAACCCCGAGAGGGGACGGAAAC-3’
Synthetic positive control (46 bp)	5’-GTTTCCGTCCCCTCTCG**ACACCATTCA**GGGTTTTGGGTCTGACGAC-3’
Reverse complement of positive control (46 bp)	5’-GTCGTCAGACCCAAAACCC**TGAATGGTGT**CGAGAGGGGACGGAAAC-3’
Biotinylated probe #1 (BP1, targets positive control)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAATCAGACCCAAAACCCC/3SpC3-3’
Biotinylated probe #2 (BP2, targets reverse complement)	5’-/52-Bio/AAAAAAAAAAAAAAAAAAAATCCGTCCCCTCTCGG/3SpC3-3’
Forward primer	5’-GTTTCCGTCCCCTCTCG-3’
Reverse primer	5’-TCGTCAGACCCAAAACCC-3’

**The LNA-substituted IS6110 reverse primer is recommended to improve specificity, but the non-LNA version is also included here because it was used for some early experiments.*

/52-Bio/ indicates dual biotin modification; /5Biosg/ indicates single biotin modification; /3SpC3/ indicates carbon spacer; “+G” and “+G” indicate locked nucleic acid (LNA) bases. Target-specific probe binding sequences are underlined. A synthetic spacer region introduced to differentiate the positive controls from endogenous Mycobacterium tuberculosis complex-specific target sequences is bolded. All DNA sequences were ordered from Integrated DNA Technologies (IDT; Coralville, IA, USA) with HPLC purification.

9.2. Sequence-specific purification protocols

9.2.1. *Surface-based hybridization capture (10 mL)*

See Section 3.4. The final hybridization protocol is also posted at http://dx.doi.org/10.17504/protocols.io.bep4jdqw.

9.2.2. *Solution-based hybridization capture (1 mL urine)*

This is the optimized solution-based hybridization capture protocol for 1 mL urine samples, described in Chapter 3 and used for comparison of urine cfDNA extraction methods in Chapter 4. This protocol was not successful in clinical samples due to the small sample volume. Directly scaling up sample volume for solution-based hybridization capture was cost prohibitive.

I. **Capture TB-specific cfDNA by hybridization**

1. Add 18.7 µL 1 µM biotinylated capture probe (final concentration 15 nM, 5'-/5BiosG/AGACCTCACCTATGTGTC/3SpC3/-3'), 224.7 µL 5M NaCl (final concentration 1 M) and 12.6 µL 1 M Tris-HCl pH 7.5 (final concentration 10 mM) to 1 mL urine in 1.5 mL DNA LoBind tube (Eppendorf). If spiking in positive control, add it now.
2. Denature in heat block at 95°C for 10 minutes.
3. Hybridize in separate heat block at 45°C for 15 minutes.

II. Immobilize capture probes on magnetic beads

1. Briefly spin down and add 83.23 µL Dynabeads MyOne Streptavidin C1 (pre-washed 3x with bead wash buffer: 1M NaCl, 10 mM Tris-HCl, 0.05% Tween) to each tube.
2. Rotate for 15 minutes at room temperature.

III. Wash to remove urine inhibitors and non-target DNA

1. Briefly spin down and place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.
2. Wash with 1 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl) by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
3. Place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.
4. Wash with 1 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl) by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
5. Place on magnetic rack for 1 minute. Remove and discard supernatant, then remove tube from magnetic rack.
6. Wash with 1 mL low salt wash buffer (15 mM NaCl, 10 mM Tris-HCl) by inverting 10-20 times, or until no bead aggregate is left on tube wall. Do not vortex. Spin down briefly.
7. Place on magnetic rack for 1 minute. Remove and discard supernatant.
8. Spin down again, place on magnetic rack, and remove as much liquid as possible using P20 pipette.

IV. Elute purified TB cfDNA

1. Add 20 µL freshly prepared 20 mM NaOH, vortex for 5 seconds, and spin down briefly.
2. Place on magnetic rack. Transfer as much supernatant as possible (usually 20–21 µL) directly to qPCR well or to new DNA LoBind tube (Eppendorf). This contains purified TB cfDNA. Avoid transferring any beads to qPCR.

3. Partially neutralize with 3.5 µL 100 mM HCl. The qPCR buffer will adjust to final pH and tolerates slightly basic pH better than acidic pH.
4. Proceed directly to qPCR.

9.2.3. *Preliminary surface-based hybridization capture protocol*

This is an early version of the surface-based hybridization capture protocol for 10 mL urine samples, prior to optimization. This protocol (with minor modifications over time) was used in the clinical pilot study described in Section 5.3. Recovery with this protocol was variable (approximately 25–60%). Major improvements from this version of the protocol to the current final version included switching from single to dual biotin modified probes, improving retention of beads during wash and elution steps, and adding a second capture probe to recover both strands of dsDNA. Minor changes include addition of beads before denaturation, denaturation at higher temperature in heat block rather than water bath, and hybridization at room temperature instead of 45°C. The final hybridization capture protocol increased detectable TB cfDNA by 4–8X (near 100% recovery of dsDNA vs 25–60% recovery of ssDNA).

I. **Immobilize capture probes on magnetic beads**

1. Vortex Dynabeads MyOne Streptavidin C1 (Thermo Fisher) for 30 seconds to ensure that beads are evenly dispersed in solution.
2. Pipette beads into a 1.5 mL DNA LoBind tube (Eppendorf). Prepare 50 µL (0.5 mg) beads per 10 mL urine sample to be analyzed.
3. Wash beads three times with an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
    i. Place beads on magnetic rack for 1 minute; remove and discard supernatant.
    ii. Add an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20), vortex for 5 seconds, and briefly spin down.
    iii. Repeat twice for a total of three washes.
4. Resuspend beads in an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
5. Add single biotinylated capture probe (5'-/5BiosG/AAAAACAGACCTCACCTATGTGT/3SpC3/-3') to beads and immediately vortex for 5 seconds. Use 50 pmol probe per 50 µL beads (0.5 µL of 100 µM stock).
6. Rotate for 15 minutes at room temperature to immobilize probes on beads.
7. Briefly spin down.

8. Wash beads three times with an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20).
    i. Place beads on magnetic rack for 1 minute; remove and discard supernatant.
    ii. Add an equal volume of high salt wash buffer (1M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20), vortex for 5 seconds, and briefly spin down.
    iii. Repeat twice for a total of three washes.
9. Resuspend beads in an equal volume high salt wash buffer.

II. **Capture TB-specific urine cfDNA by hybridization**

1. Add 2.5 mL 5 M NaCl (final concentration 1 M) to each 10 mL urine sample. If spiking in positive control, add it now.
2. Mix well by inversion.
3. Denature in water bath set to 95°C for 10 minutes (actual temp >80°C).
4. Add 127 µL 10% Tween 20 (final concentration 0.1%) and 50 µL prepared beads and gently mix.
5. Rotate for 30 minutes at 45°C in incubator to hybridize.

III. **Wash to remove urine inhibitors and non-target DNA**

1. Place on magnetic rack for 15 mL tubes and remove supernatant.
2. Add 5 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and wash by inverting until no bead aggregate is left on tube wall. Do not vortex.
3. Place on magnetic rack and remove supernatant.
4. Add 5 mL high salt wash buffer (1 M NaCl, 10 mM Tris-HCl pH 8, 0.05% Tween-20) and wash by inverting until no bead aggregate is left on tube wall. Do not vortex.
5. Place on magnetic rack and remove supernatant.
6. Add 5 mL low salt wash buffer (15 mM NaCl, 10 mM Tris-HCl pH 8, no Tween-20) and wash by inverting until no bead aggregate is left on tube wall. Do not vortex.
7. Remove all but ~1 mL wash buffer. Use remaining buffer to resuspend beads and transfer to new 1.5 mL DNA LoBind tubes.
8. Place tubes on magnetic rack for 1.5 mL tubes and remove supernatant.
9. Spin down again, place on magnetic rack, and remove as much liquid as possible using P20 pipette.

IV. **Elute purified TB cfDNA**

1. Add 20 µL freshly prepared 20 mM NaOH, vortex for 5 seconds, and spin down briefly.

2. Place on magnetic rack. Transfer as much supernatant as possible (usually 20–21 µL) directly to qPCR well or to new DNA LoBind tube (Eppendorf). This contains purified TB cfDNA. Avoid transferring any beads to qPCR.
3. Partially neutralize with 3.5 µL 100 mM HCl. The qPCR buffer will adjust to final pH and tolerates slightly basic pH better than acidic pH.
4. Proceed directly to qPCR.

9.2.4. *Surface-based hybridization with covalently coupled probes*

This protocol was used to covalently immobilize probes on beads rather than using biotin/streptavidin. It reduced the cost of beads required (from $9 to $4) but had limited recovery (~30–50%).

I. **Immobilize probes using one step EDC coupling**

1. Wash Dynabeads MyOne Carboxylic Acid (50 µL per 10 mL urine sample) twice in equal volume 100 mM MES pH 4.8.
2. Resuspend in 7.5 µL of 100 mM MES pH 4.8 per original 50 µL bead volume.
3. Prepare fresh 1.25 M EDC-HCl solution in cold 100 mM MES pH 4.8 (24 µg EDC-HCl + 100 µL MES).
4. Per 50 µL bead equivalent: Pre-mix 4 µL EDC-HCl with 900 pmol amine-modified probe (5'-/5AmMC6/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/-3'; resuspended in amine-free buffer) and additional 100 mM MES pH 4.8. Total volume should be 12.5 µL per 50 µL bead equivalent.
5. Add EDC/probe mix to beads and vortex for 10 seconds. Total coupling volume should be 20 µL per original 50 µL bead equivalent.
6. Rotate at room temperature overnight.
7. Wash 3x with TT buffer (250 mM Tris-HCL pH 7.5, 0.01% Tween-20) for 30 minutes.
8. Resuspend in TE+Tween buffer (10 mM Tris-HCl pH 7.5, 1 mM EDTA, 0.1% Tween-20).
9. Store at 4°C until use.

II. Proceed with surface-based hybridization capture (10 mL) protocol as described in Section 3.4.

9.3. PCR PROTOCOLS

9.3.1. *Conventional PCR protocol*

This protocol is used for conventional PCR of short (~40 bp) targets: IS6110, IS1081 5', IS1081 3', and DR. Primer sequences for each target are given in Table 9.1.

1. Analyze the entire eluate (approximately 24 µL) from each 10 mL urine sample in a single qPCR well. Each 50 µL reaction should contain 1.25 U OneTaq Hot Start DNA Polymerase (NEB), 1X OneTaq GC Reaction Buffer (NEB; 80 mM Tris-SO_4, 20 mM $(NH_4)_2SO_4$, 2 mM $MgSO_4$, 5% glycerol, 5% DMSO, 0.06% IGEPAL CA-630, 0.05% Tween-20, pH 9.2), 0.8 mM dNTPs (NEB), 0.4X EvaGreen (Biotium), 200 nM forward primer (Table 9.1), and 200 nM reverse primer (Table 9.1).
2. Amplify in CFX96 Touch Real-Time PCR Detection System (Bio-Rad Laboratories, Hercules, CA, USA) using an initial denaturation of 94°C for 3 min followed by 45 amplification cycles (94°C for 30s, annealing temperature [see below] for 30s, and 68°C for 1 min).

Target	PCR annealing temperature
IS6110 (LNA)*	64°C
IS6110 (no LNA)*	58°C
IS1081 (5')	58°C
IS1081 (3')	57°C
DR	56°C

**The LNA-substituted IS6110 reverse primer is recommended to improve specificity, but the non-LNA version is also included here because it was used for some early experiments.*

3. Conduct post-amplification melt analysis from 65°C to 95°C in 0.5°C increments every 5 seconds.
4. Determine C_q values at a threshold of 500 RFU, calculate recovered copies using a standard curve (0, 10, 10^2, 10^3, 10^4, and 10^5 copies positive control) run for each experiment, and verify that melting temperature (T_m) matches that of expected amplicon.

9.3.2. *Ultrashort PCR protocol*

This protocol is used for amplification of a 25 bp target within IS6110. Preliminary experiments suggest that the annealing temperature for the initial 10 cycles could be increased from 45°C to up to 57°C without reducing amplification efficiency.

1. Analyze the entire eluate (approximately 24 µL) from each 10 mL urine sample in a single qPCR well. Each 50 µL reaction should contain 1.25 U Hot Start Taq DNA Polymerase (NEB), 1X NEB Standard Taq Buffer (10 mM Tris-HCl, 50 mM KCl, 1.5 mM $MgCl_2$, pH 8.3) supplemented with an additional 0.5 mM $MgCl_2$ and 70 mM Tris-HCl, 0.8 mM dNTPs (NEB), 50 nM first stage hairpin forward primer, 700 nM second stage

universal forward primer, 700 nM reverse primer, and 100 nM MGB TaqMan probe (Table 9.1).

2. Amplify in CFX96 Touch Real-Time PCR Detection System (Bio-Rad) with an initial denaturation phase (94°C for 5 min), 10 pre-amplification cycles to extend the first stage loop primer (94°C for 30s and 45°C for 1 min), and 40 amplification cycles (94°C for 30s and 59°C for 1 min).
3. Determine C_q values at a threshold of 100 RFU and calculate recovered copies by a standard curve from 10–10^5 copies.

10. APPENDIX B: SUPPLEMENTAL INFORMATION FOR CHAPTER 3

In this appendix, I give additional assay development details about minor design considerations. During the assay development process, I tested many probe designs (different capture sequences, spacers, biotin modifications), protocol modifications (blocking steps, denaturation methods, hybridization time and temperature, wash volumes, bead volume, bead manufacturer), and other variables (Tween-20 concentration, elution pH). While these optimization experiments were individually less impactful than the "key design features" described in Chapter 3, they collectively contributed to achieving the strong performance of the final hybridization capture protocol. Many experiments in this section included few replicates because they were designed to quickly make decisions about the protocol rather than to definitively compare variables. Experiments were also conducted concurrently, so incremental improvements to the protocol during this time make direct comparison across experiments difficult.

10.1.1. *Denaturation method*

Thorough denaturation of large volume samples is essential for maximum recovery of dsDNA. Initially, I used a water bath set to 95°C for 10 minutes (actual temperature usually did not reach 95°C), which resulted in sample temperature reaching >80°C. Switching the denaturation protocol to a heat block set to 120°C for 15 minutes resulted in sample temperature reaching >90°C. The estimated T_m for the 50 bp synthetic positive control sequence in hybridization conditions (1M NaCl) is 82°C. The T_m of TB cfDNA in clinical samples may be higher for fragments longer than the target amplicon. Ensuring sample temperature of at least 85–90°C is recommended to ensure that all native TB cfDNA denatures despite the high salt concentration. I switched to using a heat block for denaturation around the same time I introduced the dual biotin modification (Section 3.7.2) and began concentrating beads to a 1 mL volume by centrifugation prior to washing (Section 10.1.6); together the three changes increased recovery from

50–60% to near 100%. Follow-up experiments confirmed that denaturation in a 95°C water bath (actual measured temperature) for 15 minutes is a suitable substitute for denaturation in a heat block of 15 mL heat block inserts are unavailable (Figure 10.1). If sample volume is to be increased >10mL, then denaturation in a boiling water bath is required (Section 7.5).

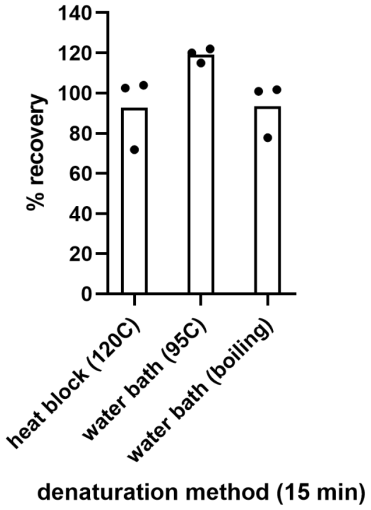


Figure 10.1: Effect of denaturation method on recovery. *Mean of n=3 shown for each method (10^3 copies 50 bp dsDNA input). Although not shown here, ≤95°C water bath for 10 min was <u>not</u> effective for denaturation.*

10.1.2. *Bead volume*

In an attempt to reduce assay cost, I evaluated whether the volume of beads used per 10 mL sample could be reduced to <50 µL. Threre was no significant decrease in recovery when using 25 µL beads (~$4.5) compared to 50 µL beads (~$9) (Figure 10.2). Recovery decreased for 10 µL or 5 µL beads. To ensure maximum recovery during early validation of the protocol, I kept 50 µL as the bead volume for now. In the future, reducing to 25 µL could be considered to reduce cost.

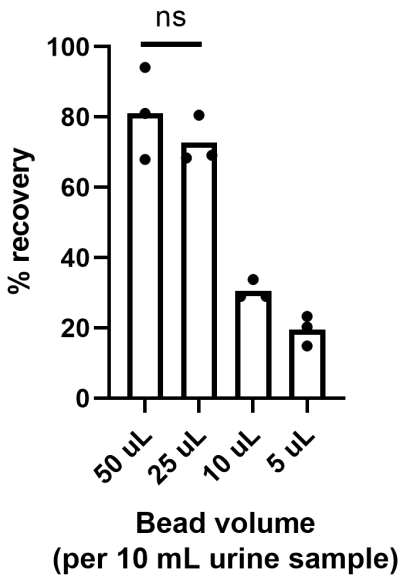


Figure 10.2: Effect of bead volume on recovery. *50, 25, 10, or 5 µL of beads were used per 10 mL urine sample (mean, n=3; 10^3 copies 50 bp ssDNA input). The difference in recovery for 50 vs 25 µL beads was not statistically significant (p=0.7, Mann-Whitney test).*

10.1.3. *Bead manufacturer*

Dynabeads MyOne Streptavidin C1 (Thermo Fisher) resulted in higher recovery than ProMag 1 Series Streptavidin (Bangs Laboratories) (Figure 10.3). Anecdotally, Dynabeads MyOne Streptavidin T1 had similar performance to Dynabeads MyOne Streptavidin C1. The surface of the C1 beads is negatively charged to minimize nonspecific binding of nucleic acids, while the T1 beads have a neutral surface.

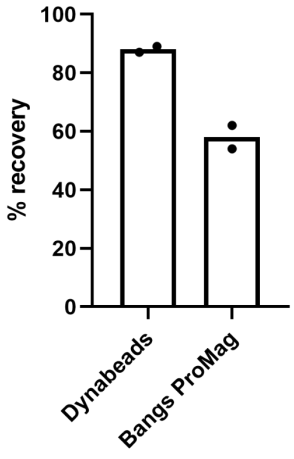


Figure 10.3: Effect of bead type on recovery. *Mean of n=2 shown for each bead type (10^3 copies 50 bp ssDNA input).*

10.1.4. *Capture probe design*

Capture probe designs I tested are listed in Table 10.1. All probe sequences are truncated versions of the PCR primer sequence. The tested probes vary in biotin modification (single or dual), biotin linker (standard C6 or TEG), and poly(A) spacer length (0, 5, or 20 bp). I tested the TEG linker, which extends the distance between the biotin and the probe sequence by 5 atoms compared to the standard C6 linker, to see if it reduced steric hindrance during hybridization. It did lead to a slight increase in recovery but was less stable in storage leading to variable performance. I returned to the standard C6 linker but added a 5 bp or 20 bp poly(A) spacer to further increase distance from the probe surface. This led to another slight increase in sensitivity to 40–60%. There was no apparent difference between the 5 bp and 20 bp poly(A) spacer for the 50 bp target, although I did not test recovery of longer fragments that would be more likely to be affected by steric hindrance. I then switched from a single biotin modification to a dual biotin modification, which reduced reliance on probe density and increased thermostabilty. The difference between single and dual biotin modification is not that apparent here (10–20% increase in recovery for dual biotin), but the difference is most noticeable in the reduced variability of the assay, likely due to improved density and distribution of probes. The dual biotin modification is discussed in detail in Section 3.7.2. I was making other simultaneous improvements to the method over the time that I designed and tested these probes, so, combined with the inherent variability of the method, it is difficult to directly attribute increases in recovery to the probe design. For example, switching to the dual biotinylated probe, combined with more effective denaturation using a heat block and introduction of a centrifuge step to concentrate beads prior to washing (1 mL wash instead of 10 mL wash) increased recovery to near 100%.

Table 10.1: Capture probes tested for surface-based hybridization.

Sequence	Biotin modification	Biotin linker	PolyA spacer	Rationale	Performance
/5BiosG/AGACCTCACCTATGTGTC/3SpC3/	Single	C6	none	Original probe design carried over from solution-based hybridization	10–15% recovery
/5BioTinTEG/CAGACCTCACCTATGTGT/3SpC3/	Single	TEG	none	Increase probe distance from bead surface to reduce steric hindrance	TEG linker less stable (30–50% recovery)
/5BioTinTEG/CAGACCTCACCTATGTGTTTTTTCAGACCTCACCTATGTGTTTTTTCAGACCTCACCTATGTGTTTTTTCAGACCTCACCTATGTGT/3SpC3/	Single	TEG	none	Include multiple target binding regions to increase binding capacity (possibly	Residual probe in PCR caused NTC amplification (no increase in recovery)

				reduce bead volume)	
/5BiosG/AAAAACAGACCTCACCTATGTGT/3SpC3/	Single	C6	5 bp	Increase probe distance from bead surface to reduce steric hindrance	Slight increase in recovery (40–60% recovery)
/5BiosG/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/	Single	C6	20 bp		Slight increase in recovery, but no more than 5 bp spacer (40–60% recovery)
/52-Bio/AAAAACAGACCTCACCTATGTGT/3SpC3/	Dual	C6	5 bp	Reduce reliance on probe density and increase thermostabilty	Reduced variation; 10–20% increase in recovery compared to single biotin modification[a]
/52-Bio/AAAAAAAAAAAAAAAAAAAACAGACCTCACCTATGTGT/3SpC3/	Dual	C6	20 bp		

[a] Dual biotinylated probes combined with switching to 1 mL wash volume (by centrifuging to concentrate beads after hybridization, see Section 10.1.6) and using a heat block instead of water bath for denaturation (Section) led to near 100% recovery.

10.1.5. *Hybridization temperature and time*

The optimal hybridization temperature would achieve both a high percent of bound target (increases with decreasing temperature) and high rate of diffusion to encourage fast probe binding (increases with increasing temperature). A general rule of thumb for hybridization is that a temperature approximately 15–20°C below the probe T_m achieves the optimal balance between these two criteria. For IS6110 capture probe T_ms of 58°C and 60°C (calculated for 2 nM probe in 1M NaCl), this corresponds to a hybridization temperature of ~45°C. Because samples must be mixed during hybridization, hybridizing above room temperature required placing the entire tube rotisserie unit within an incubator. To make processing samples easier at other sites where this may not be possible, I tested whether room temperature hybridization would result in a loss in performance. I also wanted to ensure that, regardless of temperature, near complete hybridization could occur during the duration of the incubation. Comparison of hybridization at 45°C and room temperature showed no difference in recovery and increasing the hybridization time from 30 min to 60 min did not significantly increase recovery for either temperature (Figure 10.4). 30 min hybridization at room temperature was selected for the final hybridization capture protocol.

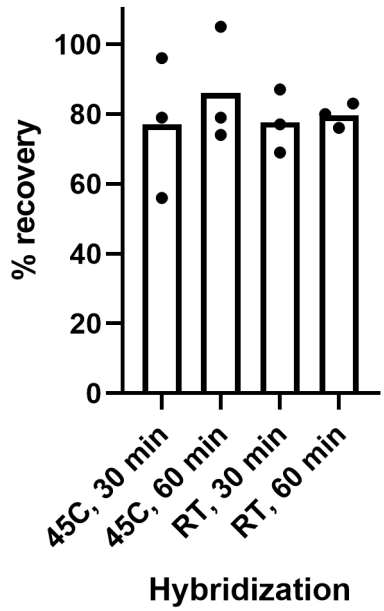


Figure 10.4: Increasing hybridization temperature (RT vs 45°C) or time (30 min vs 60 min) has no effect on recovery. *1000 copies of positive control target were extracted from 10 mL pooled urine with hybridization at room temperature or 45°C for 30 min or 60 min (mean of n=3).*

10.1.6. *Centrifugation to concentrate beads to smaller wash volume*

A small procedural change that had high impact on performance was the switch to a smaller wash volume. When I originally increased volume to 10 mL urine samples, I used a 3D-printed magnetic rack to wash the beads in 5–10 mL wash buffer (Figure 10.5). While the beads appeared to aggregate fully on the magnet, I later realized this format led to bead loss. I switched to smaller wash volumes (1 mL) by concentrating the beads by centrifugation after hybridization. This reduced bead loss, increasing recovery and reducing inter-assay variability. I made this change around the same time I introduced the dual biotin modification (Section 3.7.2) and switched to using a heat block for denaturation (Section 10.1.1); together the three changes increased recovery from 50-60% to near 100%. To avoid concentration of cellular debris along with magnetic beads, I also added a pre-hybridization centrifugation step to isolate cell-free urine at this time.

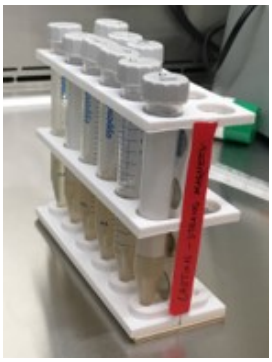


Figure 10.5: Magnetic rack for large volume wash steps.

10.1.7. *Eluate pH after neutralization*

Because the exact volume of eluate varies with residual wash buffer left on the beads, the pH after neutralization also varies. I found that the PCR tolerated high pH better than low pH, which makes sense given the slightly basic pH of the PCR buffer. To minimize effects of eluate pH variability, I made two changes. First, I only partially neutralized the eluate prior to PCR. The cfDNA is eluted in 20 µL 20 mM NaOH, which would require 4 µL 100 mM HCl for complete neutralization. Instead, I partially neutralize the eluate using 3.5 µL 100 mM HCl. Second, I use a PCR buffer with a relatively high concentration of the buffering component (80 mM Tris-SO_4) to effectively adjust to optimal pH prior to PCR.

11. APPENDIX C: SUPPLEMENTAL INFORMATION FOR CHAPTER 4

Table 11.1: Dependence of urine cfDNA extraction methods on sample pH, background DNA, and salt concentration. *For each experiment, 10^4 copies of 150 bp target were spiked into PBS (hybridization, Wizard/GuSCN, Norgen, MagMAX) or TBS (Q Sepharose) prior to extraction. Hybridization, Norgen, QIAamp, and MagMAX appear to be independent of sample composition within the expected biological range for urine. The Wizard/GuSCN method was highly dependent on urine composition, particularly pH and background DNA. The Q Sepharose method was moderately dependent on background DNA concentration in buffer but had consistent performance across biological urine replicates.*

Extraction Method	Condition	pH	Background DNA	NaCl	Percent Recovery from Buffer
Hybridization	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	126%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	117%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	116%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	116%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	113%
	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	126%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	94.7%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	97.4%

Wizard/GuSCN[a]	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	4.47%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	5.96%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	5.88%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	2.80%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	2.86%
	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	22.0%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	4.80%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	8.81%
Q Sepharose[b]	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	20.6%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	18.4%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	21.6%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	24.8%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	6.70%
	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	30.0%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	5.71%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	15.1%
Norgen	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	2.17%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	3.54%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	7.39%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	4.85%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	4.24%
	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	5.02%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	4.04%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	3.54%
QIAamp	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	15.9%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	25.8%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	15.2%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	14.1%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	12.5%
	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	25.5%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	23.8%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	16.6%
MagMAX	pH 5 (normal background DNA, normal salt)	5	100 ng/mL	137 mM	62.9%
	pH 6 (normal background DNA, normal salt)	6	100 ng/mL	137 mM	73.1%
	pH 7 (normal background DNA, normal salt)	7	100 ng/mL	137 mM	66.3%
	pH 8 (normal background DNA, normal salt)	8	100 ng/mL	137 mM	72.6%
	No background DNA (pH 6, normal salt)	6	0 ng/mL	137 mM	63.7%

	High background DNA (pH 6, normal salt)	6	1000 ng/mL	137 mM	67.7%
	Low salt (pH 6, normal background DNA)	6	100 ng/mL	13.7 mM	58.3%
	High salt (pH 6, normal background DNA)	6	100 ng/mL	500 mM	56.4%

[a] *See Figure 4.3 for results of testing Wizard/GuSCN dependence on pH and background DNA with additional technical replicates.*

[b] *See Figure 4.4 for results of testing Q Sepharose dependence on background DNA with additional technical replicates.*

12. Appendix D: Supplemental information for Chapter 5

In this appendix, I give the complete results for clinical testing described in Chapter 5.

12.1. Pilot clinical study (TB Control Program, Public Health – Seattle & King Co)

Table 12.1: Detection of urine cfDNA in 75% of pilot clinical samples from TB Control Program, Public Health – Seattle & King Co. *Urine was collected from HIV-negative patients with suspected TB and <7 days anti-TB treatment and stored up to 3 years at -80°C with 100 mM EDTA before analysis. TB urine cfDNA was extracted from 10 mL urine using an early version of the surface-based hybridization capture method (protocol in Appendix 9.2.3). Samples were classified as confident positive (detection in all replicates; dark green), weak positive (detection in single replicate or at unrealistically low concentration; light green), or negative (no TB cfDNA or incorrect melt; pink).*

Sample		Date Collected	Urine Culture Result	Treatment Status	Copies/mL detected						
					#1	#2	#3	#4	#5	**Mean**	**SD**
Healthy control		various	n/a	n/a	0	0	0	0	0	**0.00**	**0.00**
Sputum culture-negative (not TB)	UTB02	12/29/2015	-	Treatment naïve	0	0	0	n/t	n/t	**0.00**	**0.00**
	UTB03	1/19/2016	-	Day 1	0	0	n/t	n/t	n/t	**0.00**	**0.00**
	UTB04[a]	3/22/2016	-	Day 4	0.04	1.3*	0.83 (7 mL)	6.3	n/t	**0.255**	**0.331**
Sputum culture-positive	UTB01	11/24/2015	-	Day 1	0	0	0	0	n/t	**0.00**	**0.00**
	UTB05	4/5/2016	-	Treatment naïve	9.03	7.8	16.7	n/t	n/t	**1.12**	**0.482**
	UTB06	5/19/2016	-	Day 4	65.6	38.1	30	n/t	n/t	**4.46**	**1.87**
	UTB07	5/19/2016	-	Day 4	0	2.3	0	0.003	n/t	**0.058**	**0.115**
	UTB08	6/21/2016	-	Day 4	1038	561.5	n/t	n/t	n/t	**80.0**	**33.7**

	UTB09	6/24/2016	-	Day 3	225	150.3	n/t	n/t	n/t	**18.8**	**5.28**
	UTB10	7/28/2016	-	Day 4	0.38*	0.79*	0 (3 mL)	n/t	n/t	**0.059***	**0.029**
	UTB11		-	Treatment naïve	1.93	3.3	0.61	n/t	n/t	**0.195**	**0.135**
	UTB12	8/22/2016	-	Day 7	6.26	0.78	0.11	4.4	1.7	**0.265**	**0.260**
	UTB13	8/22/2016	-	Treatment naïve	1.49*	0.11*	0.02*	0	0.2 (8 mL)	**0.041***	**0.072**
	UTB14	9/1/2016	-	Treatment naïve	0.04	0.004 (6 mL)	n/t	n/t	n/t	**0.002**	**0.003**
	UTB15	9/13/2016	-	Treatment naïve	0.57	0.76	0.78	n/t	n/t	**0.070**	**0.012**

**Indicates T_m higher than that of expected amplicon (called as negative; was due to nonspecific amplification of residual human genomic DNA, which I resolved after this pilot study by redesigning the PCR primers and increasing the annealing temperature). n/t indicates not tested due to insufficient sample volume.*

Samples with insufficient volume for 10 mL have analyzed volume indicated in parentheses.

[a] Patient UTB04 had previous TB infection that may have led to detection of residual TB cfDNA.

12.2. Clinical study: Adult pulmonary TB (Edendale Hospital, South Africa)

Full results can also be found as an Excel spreadsheet in the supplemental information for the following publication:

Oreskovic A, Panpradist N, Marangu D, Ngwane MW, Magcaba ZP, Ngcobo S, Ngcobo Z, Horne DJ, Shapiro AE, Wilson DPK, Drain PK, Lutz BR. Diagnosing pulmonary tuberculosis using sequence-specific purification of urine cell-free DNA. Journal of Clinical Microbiology 59(8):e00074-21 (2021). https://doi.org/10.1128/jcm.00074-21.

Table 12.2: Demographic data, clincial data, and cfDNA detection result by participant

Reference #	Sample ID	TB status	Result by replicate									Result by participant			Copies detected		Demographic and clinical data									
			Positive defined as ≥1 copy of TB cfDNA detected AND Tm matching expected TB amplicon									Positive replicates	TB cfDNA status (+/-) Positive defined as ≥2 of 3 replicates with detected TB cfDNA	Accuracy	Mean	SD										
			#1			#2			#3																	
			Copies detected	Correct Tm?	Result (+/-)	Copies detected	Correct Tm?	Result (+/-)	Copies detected	Correct Tm?	Result (+/-)						Sex	HIV status	CD4 count	History of prior TB	Days of TB treatment	Xpert result	Culture result	Days to culture positivity	AFB smear result	Alere LAM result
1	HC01	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
2	HC02	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
3	HC03	healthy control (USA)	0.0	n/a	-	2.3	yes	+	0.0	n/a	-	1	-	true negative	0.8	1.3	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
4	HC04	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
5	HC05	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
6	HC06	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
7	HC07	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
8	HC08	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
9	HC09	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
10	HC10	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
11	HC11	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
12	HC12	healthy control (USA)	3.2	no	-	4.2	no	-	0.0	n/a	-	0	-	true negative	2.5	2.2	female	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
13	HC13	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
14	HC14	healthy control (USA)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	unknown	unknown	no	n/a	No result	No result	n/a	No result	No result
15	1083	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	1356	no	0	No result	Negative	n/a	Negative	Negative
16	1085	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	259	no	0	No result	Negative	n/a	Negative	Negative
17	1091	TB-negative control (South Africa)	0.0	n/a	-	8.2	no	-	<1	yes	-	0	-	true negative	4.1	5.8	female	+	808	no	0	No result	Negative	n/a	Negative	Negative
18	1094	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	+	766	yes	0	No result	Negative	n/a	Negative	Negative
19	1095	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	male	+	16	no	0	No result	Negative	n/a	Negative	Negative
20	1099	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	256	yes	0	No result	Negative	n/a	No result	Negative
21	1109	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	290	no	0	No result	No result	n/a	No result	Negative
22	1112	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	201	yes	0	No result	Negative	n/a	No result	Negative
23	1114	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	2.7	yes	+	1	-	true negative	0.9	1.5	female	+	345	yes	0	No result	Negative	n/a	No result	Negative
24	1118	TB-negative control (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	true negative	0.0	0.0	female	+	684	yes	0	No result	Negative	n/a	No result	Negative
25	1048	TB-positive (South Africa)	1975.0	yes	+	2718.5	yes	+	1937.6	yes	+	3	+	true positive	2210.4	440.5	female	+	545	yes	0	Positive	Positive	14	Positive (3+)	Negative
26	1049	TB-positive (South Africa)	21.1	yes	+	211.1	yes	+	35.0	yes	+	3	+	true positive	89.1	105.9	female	+	42	no	3	Positive	Positive	35	Negative	Positive (2+)
27	1050	TB-positive (South Africa)	1.6	no	-	1.7	yes	+	0.0	n/a	-	1	-	false negative	1.1	0.9	female	-	unknown	no	0	Positive	Negative	n/a	Negative	Negative
28	1051	TB-positive (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	false negative	0.0	0.0	female	-	unknown	no	0	Positive	Positive	20	Negative	Negative
29	1052	TB-positive (South Africa)	4.4	yes	+	28.1	yes	+	10.7	yes	+	3	+	true positive	14.4	12.3	female	+	475	yes	0	Positive	Positive	19	Positive (1+)	Negative
30	1053	TB-positive (South Africa)	39.1	yes	+	15.5	yes	+	63.2	yes	+	3	+	true positive	39.2	23.9	male	+	41	yes	0	Positive	Negative	n/a	Negative	Negative
31	1054	TB-positive (South Africa)	95.3	yes	+	577.7	yes	+	386.6	yes	+	3	+	true positive	353.2	242.9	male	+	42	no	2	Positive	Negative	n/a	Negative	Negative
32	1055	TB-positive (South Africa)	17090.0	yes	+	40279.6	yes	+	26763.4	yes	+	3	+	true positive	28044.3	11647.7	female	+	43	yes	0	Positive	No result	n/a	Negative	Positive (2+)
33	1056	TB-positive (South Africa)	4.7	yes	+	4.4	yes	+	214.7	yes	+	3	+	true positive	74.6	121.3	female	-	unknown	no	0	Positive	Positive	17	Negative	Negative
34	1057	TB-positive (South Africa)	399.8	yes	+	27.4	yes	+	128.4	yes	+	3	+	true positive	185.2	192.6	female	+	75	yes	0	Positive	Positive	8	Positive (1+)	Positive (2+)
35	1058	TB-positive (South Africa)	0.0	n/a	-	0.0	n/a	-	1.8	yes	+	1	-	false negative	0.6	1.0	male	+	47	yes	0	Positive	Positive	16	Negative	Negative
36	1059	TB-positive (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	false negative	0.0	0.0	female	+	577	no	0	Positive	No result	n/a	No result	Negative
37	1060	TB-positive (South Africa)	22.0	yes	+	2481.7	yes	+	771.2	yes	+	3	+	true positive	1091.6	1260.8	female	+	89	no	3	Positive	Positive	10	No result	Positive (3+)
38	1061	TB-positive (South Africa)	0.0	n/a	-	2.2	yes	+	2.0	yes	+	2	+	true positive	1.4	1.2	female	+	553	no	0	Positive	Negative	n/a	Negative	Negative
39	1062	TB-positive (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	false negative	0.0	0.0	female	-	unknown	no	0	Positive	No result	n/a	Negative	Negative
40	1064	TB-positive (South Africa)	0.0	n/a	-	0.0	n/a	-	0.0	n/a	-	0	-	false negative	0.0	0.0	male	+	499	yes	0	Positive	Positive	14	Negative	Negative
41	1066	TB-positive (South Africa)	2226.7	yes	+	4434.5	yes	+	3689.1	yes	+	3	+	true positive	3450.1	1123.1	male	+	18	yes	0	Positive	Positive	8	Negative	Positive (2+)
42	1067	TB-positive (South Africa)	658.0	yes	+	808.7	yes	+	921.9	yes	+	3	+	true positive	796.2	132.4	male	+	61	no	3	Positive	Positive	16	No result	Positive (3+)
43	1068	TB-positive (South Africa)	5831.6	yes	+	3653.1	yes	+	2954.7	yes	+	3	+	true positive	4146.5	1500.5	male	-	unknown	yes	0	Positive	Positive	5	Positive (3+)	Negative
44	1069	TB-positive (South Africa)	17.3	yes	+	13.1	yes	+	0.0	n/a	-	2	+	true positive	10.1	9.0	male	-	unknown	yes	0	Positive	Positive	10	Positive (2+)	Negative
45	1070	TB-positive (South Africa)	<1	yes	-	0.0	n/a	-	3.2	yes	+	1	-	false negative	1.6	2.2	male	-	unknown	no	0	Positive	Positive	8	Negative	Negative
46	1071	TB-positive (South Africa)	45.6	yes	+	24.5	yes	+	46.9	yes	+	3	+	true positive	39.0	12.6	male	-	unknown	no	0	Positive	No result	n/a	Positive (3+)	Negative
47	1072	TB-positive (South Africa)	387.4	yes	+	351.1	yes	+	380.5	yes	+	3	+	true positive	373.0	19.2	male	-	unknown	no	0	Positive	Positive	17	Negative	Negative
48	1073	TB-positive (South Africa)	444.8	yes	+	819.5	yes	+	465.6	yes	+	3	+	true positive	576.6	210.6	male	+	18	yes	0	Positive	Positive	15	Negative	Positive (1+)
49	1074	TB-positive (South Africa)	6.0	yes	+	3.1	yes	+	2.7	yes	+	3	+	true positive	3.9	1.8	female	+	161	no	3	Positive	Negative	n/a	Negative	Negative
50	1075	TB-positive (South Africa)	43.7	yes	+	88.5	yes	+	351.7	yes	+	3	+	true positive	161.3	166.4	female	+	101	no	3	Positive	Negative	n/a	Negative	Negative
51	1076	TB-positive (South Africa)	0.9	yes	-	5.0	yes	+	7.9	yes	+	2	+	true positive	4.6	3.5	female	+	350	no	0	Positive	Positive	8	Positive (2+)	Negative
52	1078	TB-positive (South Africa)	45.2	yes	+	1150.9	yes	+	293.0	yes	+	3	+	true positive	496.3	580.3	male	-	unknown	no	0	Positive	Positive	5	Positive (3+)	Negative
53	1079	TB-positive (South Africa)	392.3	yes	+	1754.4	yes	+	1966.8	yes	+	3	+	true positive	1371.2	854.4	male	+	2	no	3	Positive	Positive	15	Negative	Positive (3+)
54	1080	TB-positive (South Africa)	2.2	yes	+	0.0	n/a	-	<1	yes	-	1	-	false negative	1.1	1.6	male	+	102	no	0	Positive	Negative	n/a	No result	Negative
55	1081	TB-positive (South Africa)	218.8	yes	+	329.7	yes	+	161.1	yes	+	3	+	true positive	236.5	85.7	male	+	24	no	0	Positive	Positive	n/a	Positive (1+)	Negative
56	1082	TB-positive (South Africa)	139.3	yes	+	235.5	yes	+	38.3	yes	+	3	+	true positive	137.7	98.6	female	+	329	yes	0	Positive	Positive	6	Positive (3+)	Positive (1+)
57	1084	TB-positive (South Africa)	2.3	yes	+	8.2	yes	+	1.1	yes	+	3	+	true positive	3.9	3.8	male	+	81	yes	0	Positive	No result	n/a	Negative	Negative
58	1086	TB-positive (South Africa)	956.9	yes	+	1100.3	yes	+	954.5	yes	+	3	+	true positive	1003.9	83.5	male	-	unknown	yes	0	Positive	Negative	n/a	Negative	Negative
59	1087	TB-positive (South Africa)	14.4	yes	+	27.2	yes	+	9.6	yes	+	3	+	true positive	17.1	9.1	male	-	unknown	no	0	Positive	Positive	22	Negative	Negative
60	1089	TB-positive (South Africa)	1.7	yes	+	3.1	yes	+	10.3	yes	+	3	+	true positive	5.0	4.6	female	+	222	no	0	Positive	Positive	22	Positive (3+)	Negative
61	1090	TB-positive (South Africa)	4.6	yes	+	0.0	n/a	-	3.5	yes	+	2	+	true positive	2.7	2.4	male	+	380	no	0	Positive	Positive	22	Negative	Negative
62	1092	TB-positive (South Africa)	17980.0	yes	+	18839.7	yes	+	18783.4	yes	+	3	+	true positive	18534.4	480.9	male	+	5	no	1	Positive	No result	n/a	Positive (1+)	Positive (2+)
63	1093	TB-positive (South Africa)	7.8	yes	+	4.8	yes	+	7.5	yes	+	3	+	true positive	6.7	1.6	male	-	unknown	yes	0	Positive	Positive	14	Positive (3+)	Negative
64	1101	TB-positive (South Africa)	15.5	yes	+	20.1	yes	+	80.4	yes	+	3	+	true positive	38.7	36.2	female	+	78	no	0	Positive	Positive	8	Negative	Negative
65	1103	TB-positive (South Africa)	5257.5	yes	+	5320.6	yes	+	4484.0	yes	+	3	+	true positive	5020.7	465.9	male	-	unknown	no	2	Positive	Positive	8	Positive (3+)	Negative
66	1105	TB-positive (South Africa)	3590.6	yes	+	2004.6	yes	+	1414.4	yes	+	3	+	true positive	2336.5	1125.4	female	+	17	no	0	Positive	No result	n/a	sitive (scant	Positive (1+)
67	1106	TB-positive (South Africa)	5.0	yes	+	0.0	n/a	-	6.5	yes	+	2	+	true positive	3.8	3.4	female	-	unknown	yes	0	Positive	Positive	5	No result	Negative
68	1107	TB-positive (South Africa)	20.5	yes	+	71.9	yes	+	45.4	yes	+	3	+	true positive	45.9	25.7	male	+	58	no	1	Positive	Positive	13	Negative	Positive (1+)
69	1110	TB-positive (South Africa)	76.7	yes	+	118.1	yes	+	161.0	yes	+	3	+	true positive	118.6	42.2	female	+	115	no	2	Positive	Positive	9	Positive (1+)	Positive (3+)
70	1115	TB-positive (South Africa)	165.3	yes	+	165.5	yes	+	106.3	yes	+	3	+	true positive	145.7	34.1	male	+	265	no	0	Positive	Positive	4	Positive (3+)	Negative
71	1116	TB-positive (South Africa)	1950.0	yes	+	3100.1	yes	+	3504.1	yes	+	3	+	true positive	2851.4	806.3	male	+	653	no	1	Positive	Positive	4	Positive (3+)	Positive (1+)
72	1117	TB-positive (South Africa)	407.0	yes	+	1634.9	yes	+	1886.8	yes	+	3	+	true positive	1309.5	791.7	male	+	455	no	1	Positive	Positive	n/a	Negative	Negative
73	1119	TB-positive (South Africa)	15.4	yes	+	24.8	yes	+	43.6	yes	+	3	+	true positive	27.9	14.4	female	+	29	no	0	Positive	Positive	17	Positive (3+)	Positive (3+)

Table 12.3: Urine characteristics, bead behavior problems, and cfDNA detection by participant

Reference #	Sample ID	TB status	Accuracy	Urine characteristics (color on a subjective scale of 1 = very light, 2 = light, 3 = medium, 4 = dark; others read visually using Fisherbrand 10-SG Urine Reagent Strips)										
				color	glucose	billirubin	ketone	specific gravity	blood	pH	protein	urobilinogen	nitrite	leukocytes
1	HC01	healthy control (USA)	true negative	3.5	-	+	-	1.025	-	6.0	-	0.2	-	-
2	HC02	healthy control (USA)	true negative	2	-	-	-	1.010	-	7.5	-	0.2	-	-
3	HC03	healthy control (USA)	true negative	2	-	-	-	1.000	-	9.0	-	0.2	-	-
4	HC04	healthy control (USA)	true negative	2	-	-	-	1.005	-	7.5	-	0.2	-	-
5	HC05	healthy control (USA)	true negative	2.5	-	-	-	1.010	-	6.5	-	0.2	-	-
6	HC06	healthy control (USA)	true negative	2	-	-	-	1.000	-	9.0	-	0.2	-	-
7	HC07	healthy control (USA)	true negative	2.5	-	+	-	1.010	-	6.5	-	0.2	-	-
8	HC08	healthy control (USA)	true negative	2	-	-	-	1.000	-	9.0	-	0.2	-	-
9	HC09	healthy control (USA)	true negative	2	-	-	-	1.000	-	8.0	-	0.2	-	-
10	HC10	healthy control (USA)	true negative	2	-	-	-	1.000	-	8.0	-	0.2	-	-
11	HC11	healthy control (USA)	true negative	2	-	-	-	1.000	-	8.0	-	0.2	-	-
12	HC12	healthy control (USA)	true negative	2	-	-	-	1.000	-	8.0	-	0.2	-	-
13	HC13	healthy control (USA)	true negative	2	-	-	-	1.000	+/-	8.0	-	0.2	-	-
14	HC14	healthy control (USA)	true negative	2	-	-	-	1.000	-	8.0	-	0.2	-	-
15	1083	TB-negative control (South Africa)	true negative	2.5	-	-	-	1.020	+	5.0	+	0.2	-	+
16	1085	TB-negative control (South Africa)	true negative	2.5	-	-	-	1.005	-	7.5	+	0.2	-	-
17	1091	TB-negative control (South Africa)	true negative	2	-	+	-	1.000	-	9.0	++	0.2	-	-
18	1094	TB-negative control (South Africa)	true negative	2	-	-	-	1.005	-	6.5	+/-	0.2	-	-
19	1095	TB-negative control (South Africa)	true negative	2	-	-	-	1.005	+++	7.5	++	0.2	-	-
20	1099	TB-negative control (South Africa)	true negative	3	-	-	-	1.015	++	6.5	+	0.2	-	-
21	1109	TB-negative control (South Africa)	true negative	2	-	-	-	1.015	-	6.5	+/-	0.2	-	-
22	1112	TB-negative control (South Africa)	true negative	2	-	-	-	1.005	-	8.0	+/-	0.2	-	-
23	1114	TB-negative control (South Africa)	true negative	2	-	-	-	1.005	+/-	8.0	+/-	0.2	-	++
24	1118	TB-negative control (South Africa)	true negative	1.5	-	-	-	1.010	-	7.0	+	≤1	-	-
25	1048	TB-positive (South Africa)	true positive	2.5	-	-	-	1.020	+++	6.5	++	0.2	-	-
26	1049	TB-positive (South Africa)	true positive	2.5	-	-	-	1.015	++	6.5	++	0.2	-	-
27	1050	TB-positive (South Africa)	false negative	2	-	-	-	1.010	-	7.0	+/-	0.2	-	-
28	1051	TB-positive (South Africa)	false negative	3	-	+	-	1.010	-	6.0	+	1	-	-
29	1052	TB-positive (South Africa)	true positive	3	-	-	-	1.010	-	6.0	+	0.2	-	-
30	1053	TB-positive (South Africa)	true positive	3	+/-	+	-	1.025	-	6.0	+++	0.2	-	-
31	1054	TB-positive (South Africa)	true positive	3.5	-	+	-	1.020	-	6.0	++	0.2	-	-
32	1055	TB-positive (South Africa)	true positive	3	-	-	-	1.015	++	6.0	++	0.2	-	-
33	1056	TB-positive (South Africa)	true positive	2.5	-	-	-	1.015	-	6.0	+/-	0.2	-	+
34	1057	TB-positive (South Africa)	true positive	3.5	-	-	-	1.015	++	6.0	++	0.2	-	-
35	1058	TB-positive (South Africa)	false negative	2.5	-	-	-	1.015	-	6.0	+/-	0.2	-	-
36	1059	TB-positive (South Africa)	false negative	2.5	-	-	-	1.020	+	6.0	+/-	0.2	-	-
37	1060	TB-positive (South Africa)	true positive	3	-	-	-	1.005	-	7.5	+/-	0.2	-	-
38	1061	TB-positive (South Africa)	true positive	3	-	-	-	1.010	-	6.5	+/-	0.2	-	-
39	1062	TB-positive (South Africa)	false negative	2.5	-	-	-	1.010	+	7.5	+/-	0.2	-	-
40	1064	TB-positive (South Africa)	false negative	2.5	-	-	-	1.010	-	7.5	+/-	0.2	-	-
41	1066	TB-positive (South Africa)	true positive	2.5	-	-	-	1.010	-	6.0	+	0.2	-	-
42	1067	TB-positive (South Africa)	true positive	3	-	+	-	1.020	-	6.0	++	0.2	-	-
43	1068	TB-positive (South Africa)	true positive	3	-	-	-	1.010	-	6.5	+/-	0.2	-	-
44	1069	TB-positive (South Africa)	true positive	3	-	-	-	1.015	-	6.5	++	0.2	-	-
45	1070	TB-positive (South Africa)	false negative	1	-	-	-	1.000	-	7.5	+/-	0.2	-	-
46	1071	TB-positive (South Africa)	true positive	2.5	-	-	-	1.025	-	5.0	+/-	0.2	-	-
47	1072	TB-positive (South Africa)	true positive	2.5	-	-	-	1.005	-	7.5	+/-	0.2	-	-
48	1073	TB-positive (South Africa)	true positive	3	-	+	-	1.020	-	6.0	+++	0.2	-	-
49	1074	TB-positive (South Africa)	true positive	2.5	-	-	-	1.010	-	7.0	+	0.2	-	-
50	1075	TB-positive (South Africa)	true positive	3.5	-	+	-	1.020	-	8.0	+++	0.2	-	+++
51	1076	TB-positive (South Africa)	true positive	2.5	-	+	-	1.010	-	7.0	+	0.2	-	-
52	1078	TB-positive (South Africa)	true positive	2.5	-	+	-	1.015	-	5.0	+	0.2	-	-
53	1079	TB-positive (South Africa)	true positive	2.5	-	+	-	1.015	-	5.0	+	0.2	-	-
54	1080	TB-positive (South Africa)	false negative	2.5	-	-	-	1.010	-	7.5	+	0.2	-	-
55	1081	TB-positive (South Africa)	true positive	2	-	-	-	1.010	-	6.0	+	0.2	-	-
56	1082	TB-positive (South Africa)	true positive	2.5	-	-	-	1.005	-	6.5	+	0.2	-	-
57	1084	TB-positive (South Africa)	true positive	3	-	+	-	1.025	+++	5.0	++	0.2	-	-
58	1086	TB-positive (South Africa)	true positive	2.5	-	-	-	1.005	-	7.5	+	0.2	-	++
59	1087	TB-positive (South Africa)	true positive	3	-	-	-	1.010	-	7.0	+/-	0.2	-	-
60	1089	TB-positive (South Africa)	true positive	2	-	-	-	1.005	++	6.5	+/-	0.2	-	++
61	1090	TB-positive (South Africa)	true positive	3	-	-	-	1.015	+	6.5	+	0.2	-	-
62	1092	TB-positive (South Africa)	true positive	3	-	++	-	1.020	-	6.0	+	2	-	-
63	1093	TB-positive (South Africa)	true positive	2	-	-	-	1.010	-	6.5	+/-	0.2	-	-
64	1101	TB-positive (South Africa)	true positive	2.5	-	-	-	1.015	+++	6.5	+	0.2	-	-
65	1103	TB-positive (South Africa)	true positive	3.5	++++	-	-	1.010	-	7.5	+	0.2	-	-
66	1105	TB-positive (South Africa)	true positive	3	-	-	-	1.010	-	6.5	++	0.2	-	-
67	1106	TB-positive (South Africa)	true positive	2.5	-	-	-	1.005	-	8.0	+	0.2	-	-
68	1107	TB-positive (South Africa)	true positive	2.5	-	-	-	1.005	+++	6.5	++	0.2	-	-
69	1110	TB-positive (South Africa)	true positive	3.5	-	+	-	1.015	+	6.0	+	0.2	-	++
70	1115	TB-positive (South Africa)	true positive	2.5	-	-	-	1.010	-	7.5	+/-	0.2	-	-
71	1116	TB-positive (South Africa)	true positive	4	-	-	-	1.020	-	6.0	+	0.2	-	-
72	1117	TB-positive (South Africa)	true positive	2.5	-	-	-	1.015	-	6.5	++	≤1	-	-
73	1119	TB-positive (South Africa)	true positive	3	-	-	-	1.015	+++	6.5	++	≤1	-	-

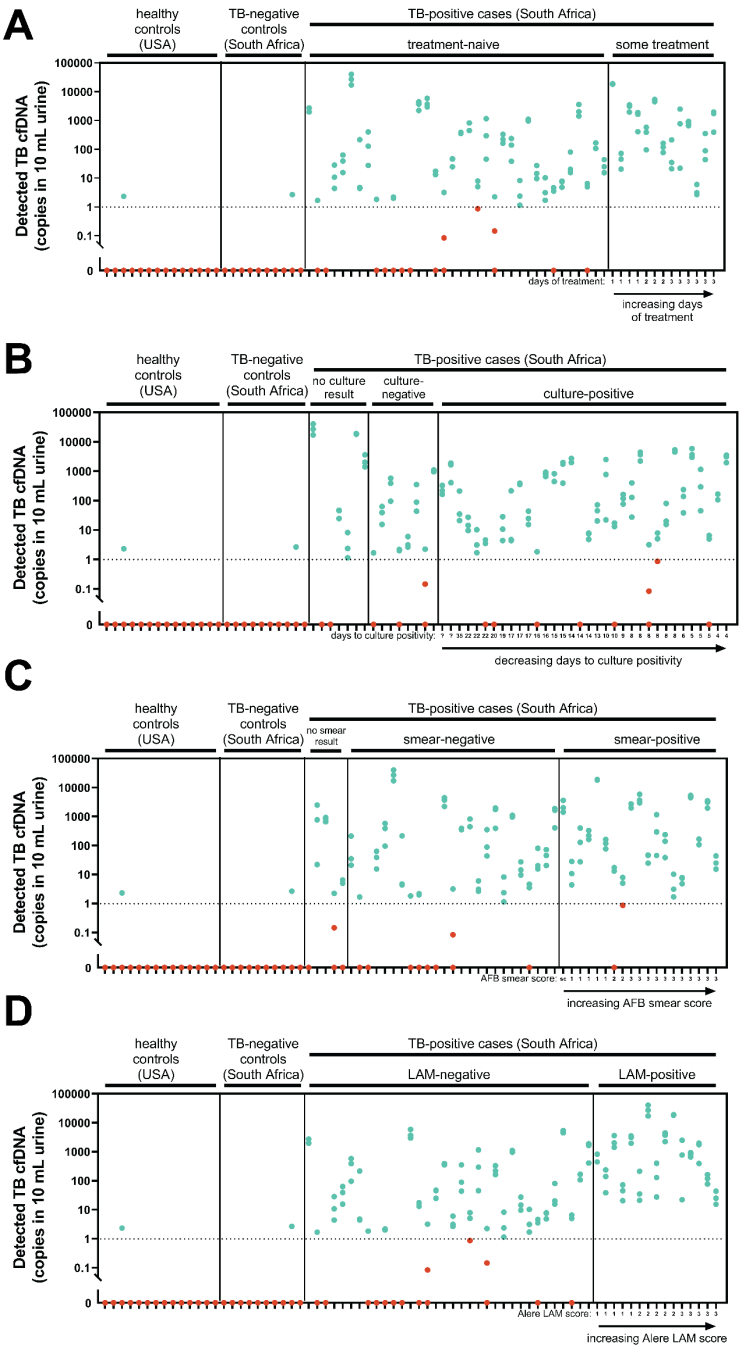


Figure 12.1: Additional correlations with detected concentrations of TB-specific urine cfDNA. *For all plots, each dot represents one of three replicates per sample, processed on different days. Cyan dots represent positive replicates with ≥1 copy detected and melting temperature matching the expected IS6110 amplicon. Red dots represent negative replicates. At least two out of three replicates were required to be positive for a sample to be called as positive.* ***(A)*** *Stratified by treatment status and ranked by days of treatment (no correlation).* ***(B)*** *Stratified by culture result and ranked by days to culture positivity (Spearman's ρ = -0.36 [95% CI: -0.64 to -0.0060], p = 0.041).* ***(C)*** *Stratified by smear result and ranked by smear score (no correlation).* ***(D)*** *Stratified by Alere urine LAM result and ranked by Alere urine LAM score (no correlation).*

Table 12.4: Diagnostic odds ratios indicating associations with a positive urine cfDNA result.

Variable	Comparison	Odds ratio (95% CI)[a]	P-value[b]
HIV status	positive vs negative	2.7 (0.68 – 10.5)	0.23
CD4 count[c]	≤200 vs >200 cells/mm^3	2.0 (0.28 – 13.9)	0.60
TB treatment status	treatment-naïve vs some treatment	infinity (0.79 – infinity)	0.090
Sputum culture result	positive vs negative	2.5 (0.39 – 14.7)	0.32
AFB sputum smear result	positive vs negative	infinity (1.6 – infinity)	0.029*
Alere urine LAM result	positive vs negative	infinity (1.0 – infinity)	0.087

** Indicates P < 0.05*

[a] Diagnostic odds ratio of >1 indicates variables associated with higher likelihood of a positive cfDNA result. 95% confidence intervals calculated using the Baptista-Pike method.

[b] Diagnostic odds ratios were compared to a value of 1 using Fisher's exact test.

[c] CD4 count was measured for HIV-positive patients only.

Table 12.5: Correlations with detected TB-specific cfDNA concentration.

Variable	Spearman's rank correlation coefficient (95% CI)	P-value
CD4 count[a]	-0.43 (-0.68 to -0.10)	0.011*
Days of TB treatment[b]	-0.36 (-0.77 to 0.26)	0.23
Days to sputum culture positivity	-0.36 (-0.64 to -0.0060)	0.041*
AFB sputum smear score[c]	-0.37 (-0.77 to 0.24)	0.21
Alere urine LAM score[d]	-0.094 (-0.59 to 0.45)	0.76

** Indicates P < 0.05*

[a] CD4 count was measured for HIV-positive patients only.

[b] Correlation calculated only for patients with some treatment (1 – 3 days). Treatment-naïve patients were excluded.

[c] Correlation calculated only for smear-positive patients (AFB score ≥1). Smear-negative patients were excluded.

[d] Correlation calculated only for LAM-positive patients (Alere LAM score ≥1). LAM-negative patients were excluded.

Table 12.6: Comparison of detected TB-specific cfDNA concentration across groups.

Variable	Groups for comparison of detected cfDNA concentration	P-value[a]
HIV status	positive vs negative	0.8007
CD4 count[b]	≤200 vs >200 cells/mm^3	0.1946
TB treatment status	treatment-naïve vs some treatment	0.0447*
Sputum culture result	positive vs negative	0.3709
AFB sputum smear result	positive vs negative	0.7013
Alere urine LAM result	positive vs negative	0.0045**
Gender	female vs male	0.1081

** Indicates P < 0.05; ** Indicates P < 0.01*

[a] P-values calculated using Mann-Whitney test comparing sample means of cfDNA-positive samples.

[b] CD4 count was measured for HIV-positive patients only.

13. APPENDIX E: SUPPLEMENTAL INFORMATION FOR CHAPTER 6

In this appendix, I give the complete results for NGS of urine cfDNA described in Chapter 6.

Full results can also be found as Excel spreadsheets in the supplemental information for the following publication:

Unbiased sequencing of *Mycobacterium tuberculosis* urinary cell-free DNA reveals extremely short fragment lengths. Oreskovic A, Waalkes A, Holmes EA, Rosenthal CA, Wilson DPK, Shapiro AE, Drain PK, Lutz BR, Salipante SJ (manuscript in preparation for submission to Journal of Clinical Microbiology).

Table 13.1: Demographic and clincial data by participant

Participant ID	Sample ID	TB status (Xpert MTB/RIF Ultra)	TB cfDNA qPCR result (IS6110)	Demographic and clinical data						
				Gender	HIV status	CD4 count	Days on TB treatment	Culture result	Days to culture positivity	Alere LAM result
01	016	TB-positive	cfDNA-positive	Male	HIV-positive	82	0	culture-negative	n/a	Negative
02	052	TB-positive	cfDNA-positive	Female	HIV-positive	623	0	culture-positive	16	Negative
03	071	TB-positive	cfDNA-positive	Male	HIV-positive	18	0	culture-positive	8	Positive (4+)
04	101	TB-positive	cfDNA-positive	Male	HIV-positive	21	0	culture-negative	n/a	Positive (3+)
05	112	TB-positive	cfDNA-positive	Female	HIV-positive	43	0	culture-positive	12	Positive (2+)
06	137	TB-positive	cfDNA-positive	Male	HIV-positive	141	1	culture-negative	n/a	Positive (2+)
07	138	TB-positive	cfDNA-positive	Male	HIV-positive	249	0	culture-positive	6	Negative
08	161	TB-positive	cfDNA-positive	Female	HIV-positive	75	0	culture-positive	12	Positive (2+)
09	260	TB-positive	cfDNA-positive	Female	HIV-positive	unknown	0	culture-positive	7	Negative
10	008	TB-positive	cfDNA-negative	Male	HIV-positive	783	0	culture-negative	n/a	Negative
11	005	TB-negative	cfDNA-negative	Male	HIV-positive	648	0	culture-negative	n/a	Negative
12	006	TB-negative	cfDNA-negative	Female	HIV-positive	409	0	culture-negative	n/a	Negative

Table 13.2: Sequencing library metrics by participant

Participant ID	Sample ID	TB status (Xpert MTB/RIF Ultra)	TB cfDNA qPCR result (IS6110)	Sequencing library metrics			metagenomic classification analysis (kraken2)						mapping-based analysis			
				Total reads	Final reads	Library complexity	Number of fragments classified as bacteria	% of fragments classified as bacteria	Number of fragments classified as TB	% of fragments classified as TB (after mapping)	Number of fragments classified as Homo sapiens	% of fragments classified as Homo sapiens	Number of fragments mapped to TB	% of fragments mapped to TB	Number of fragments mapped to human	% of fragments mapped to human
01	016	TB-positive	cfDNA-positive	101760688	89531271	0.983	663623	0.741219233	61	6.8133E-05	74291145	82.97787373	667	0.000744991	88426571	98.76612943
02	052	TB-positive	cfDNA-positive	77245725	63353734	0.966	2339531	3.692806804	7	1.1049E-05	58977843	93.0929233	38250	0.060375289	61879693	97.6733163
03	071	TB-positive	cfDNA-positive	113742998	97162514	0.926	555380	0.571599043	19547	2.0118E-02	88142654	90.71672847	78240	0.080524882	96637068	99.45920913
04	101	TB-positive	cfDNA-positive	58587870	46041712	0.955	1161042	2.521717698	145	3.1493E-04	45684934	99.22509832	7784	0.016906409	45684934	99.22509832
05	112	TB-positive	cfDNA-positive	64186517	50359835	0.981	1346265	2.673291126	118	2.3431E-04	47619646	94.55878082	2281	0.004529403	49923505	99.1335754
06	137	TB-positive	cfDNA-positive	102162546	87190778	0.952	1212771	1.390939533	2184	2.5049E-03	73679886	84.50421901	26964	0.030925289	86069747	98.71427802
07	138	TB-positive	cfDNA-positive	82446286	67584770	0.984	653814	0.967398424	54	7.9900E-05	63648486	94.1757825	255	0.000377304	66373618	98.20795129
08	161	TB-positive	cfDNA-positive	109109767	99091322	0.969	1156416	1.167020458	1078	1.0879E-03	93892556	94.75356076	22222	0.022425778	98213009	99.11363278
09	260	TB-positive	cfDNA-positive	106467683	93679790	0.969	1553153	1.657938174	126	1.3450E-04	88549265	94.52333849	48786	0.052077401	92270660	98.4958015
10	008	TB-positive	cfDNA-negative	50209616	40010907	1.000	423117	1.057504145	4	9.9973E-06	38140481	95.3252097	4	9.99727E-06	39774497	99.40913611
11	005	TB-negative	cfDNA-negative	34496174	28263144	1.000	793799	2.808601195	0	0.0000E+00	26946143	95.34021763	9	3.18436E-05	28086587	99.37531012
12	006	TB-negative	cfDNA-negative	29951750	23576658	0.974	289050	1.226000733	0	0.0000E+00	21608917	91.65385951	42	0.000178142	23359736	99.07992897

Table 13.3: cfDNA characterization results by participant

Participant ID	Sample ID	TB status (Xpert MTB/RIF Ultra)	TB cfDNA qPCR result (IS6110)	cfDNA concentration				cfDNA fragment length (by mapping-based analysis)				cfDNA fragment length (by kraken2 metagenomic classification analysis)			
				TB cfDNA in eluate (IS6110 copies/uL)	TB cfDNA in urine (IS6110 copies in 10 mL urine, calculated)	Total cfDNA in eluate (ng/uL)	Total cfDNA in urine (ng/mL, calculated)	mode human fragment length (bp)	mode TB fragment length(s) (bp)	median (IQR) human fragment length (bp)	median (IQR) TB fragment length (bp)	mode human fragment length (bp)	mode TB fragment length(s) (bp)	median (IQR) human fragment length (bp)	median (IQR) TB fragment length (bp)
01	016	TB-positive	cfDNA-positive	5.7	283	24.2	121	38	19	59 (42-106)	21 (19-41)	28	52, 61, 432*	67 (42-127)	76 (50-171)*
02	052	TB-positive	cfDNA-positive	2.1	106	9.06	45.3	53	19	92 (60-160)	19 (19-21)	53	40, 97, 105*	97 (63-176)	97 (40-105)*
03	071	TB-positive	cfDNA-positive	792	39575	7.98	39.9	29	19	42 (37-50)	22 (20-30)	29	38	45 (38-65)	44 (39-50)
04	101	TB-positive	cfDNA-positive	20.2	1010	5.00	25	29	19	89 (51-144)	19 (19-20)	29	39	95 (54-154)	39 (36-43)
05	112	TB-positive	cfDNA-positive	21.2	1062	2.14	10.7	41	19	67 (47-106)	19 (19-20)	41	43	72 (49-130)	42 (37-63)
06	137	TB-positive	cfDNA-positive	358	17893	2.8	14	38	19	53 (40-95)	19 (19-21)	38	39, 43	69 (42-135)	61 (44-106)
07	138	TB-positive	cfDNA-positive	5.4	270	2.9	14.5	42	32	86 (54-136)	45 (36-58)	41	38	93 (57-152)	39 (36-44)
08	161	TB-positive	cfDNA-positive	14.8	738	39.2	196	41	19	73 (49-147)	19 (19-20)	40	39	82 (51-171)	45 (40-52)
09	260	TB-positive	cfDNA-positive	6.8	338	4.2	21	41	19	85 (53-142)	19 (19-20)	42	38	92 (55-155)	39 (37-45)
10	008	TB-positive	cfDNA-negative	0	0	30.0	150	41	35, 42, 61, 70*	92 (56-148)	52 (39-66)*	42	61, 71*	93 (57-149)	66 (61-70)*
11	005	TB-negative	cfDNA-negative	0	0	7.04	35.2	42	n/a	73 (51-106)	n/a	42	n/a	78 (52-123)	n/a
12	006	TB-negative	cfDNA-negative	0	0	9.28	46.4	41	n/a	73 (49-119)	n/a	41	n/a	73 (50-124)	n/a
								* indicates samples with low TB read count, which may have hindered accurate determination of mode, median, and IQR fragment lengths							

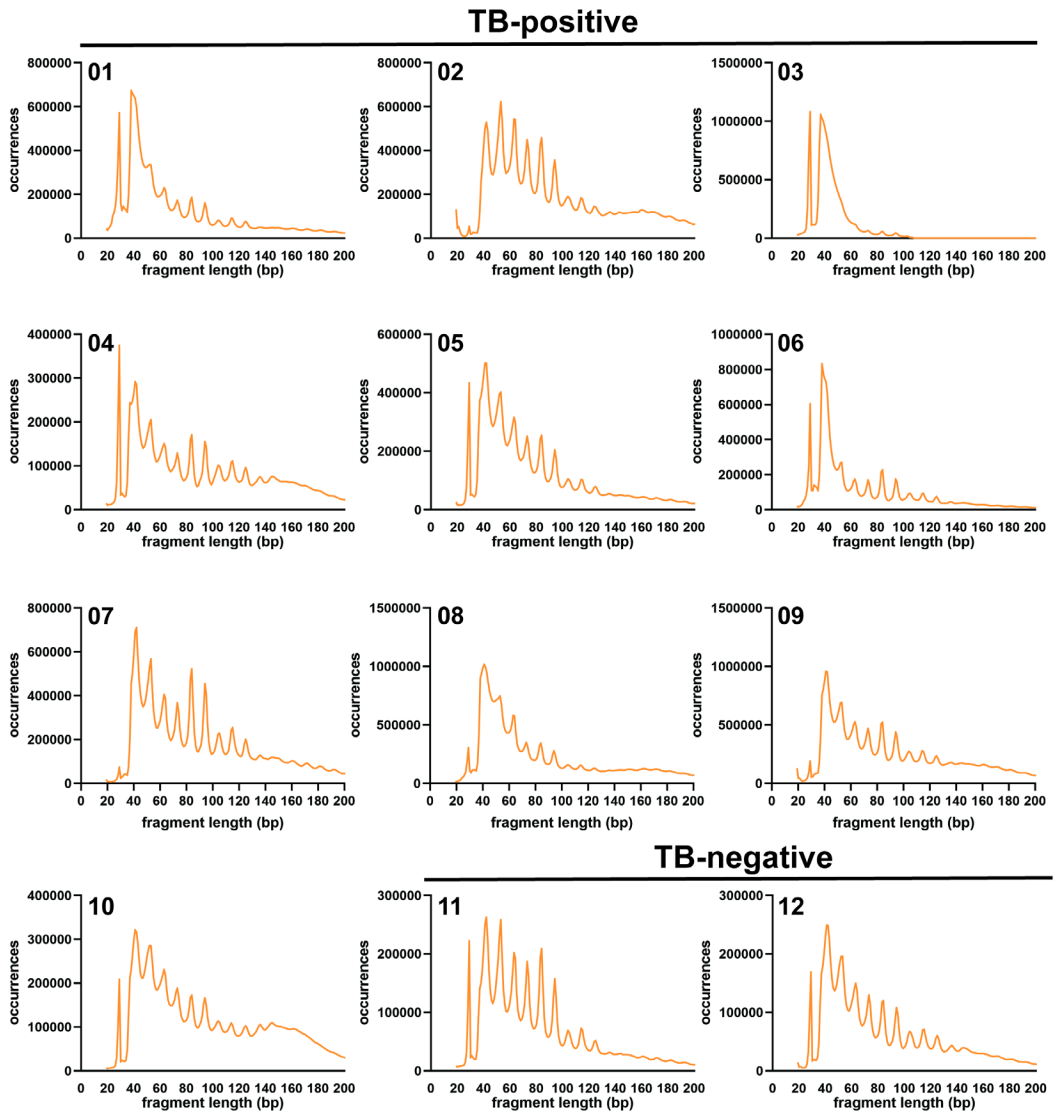


Figure 13.1: Fragment length distribution of cfDNA mapped to the human genome in each sample.

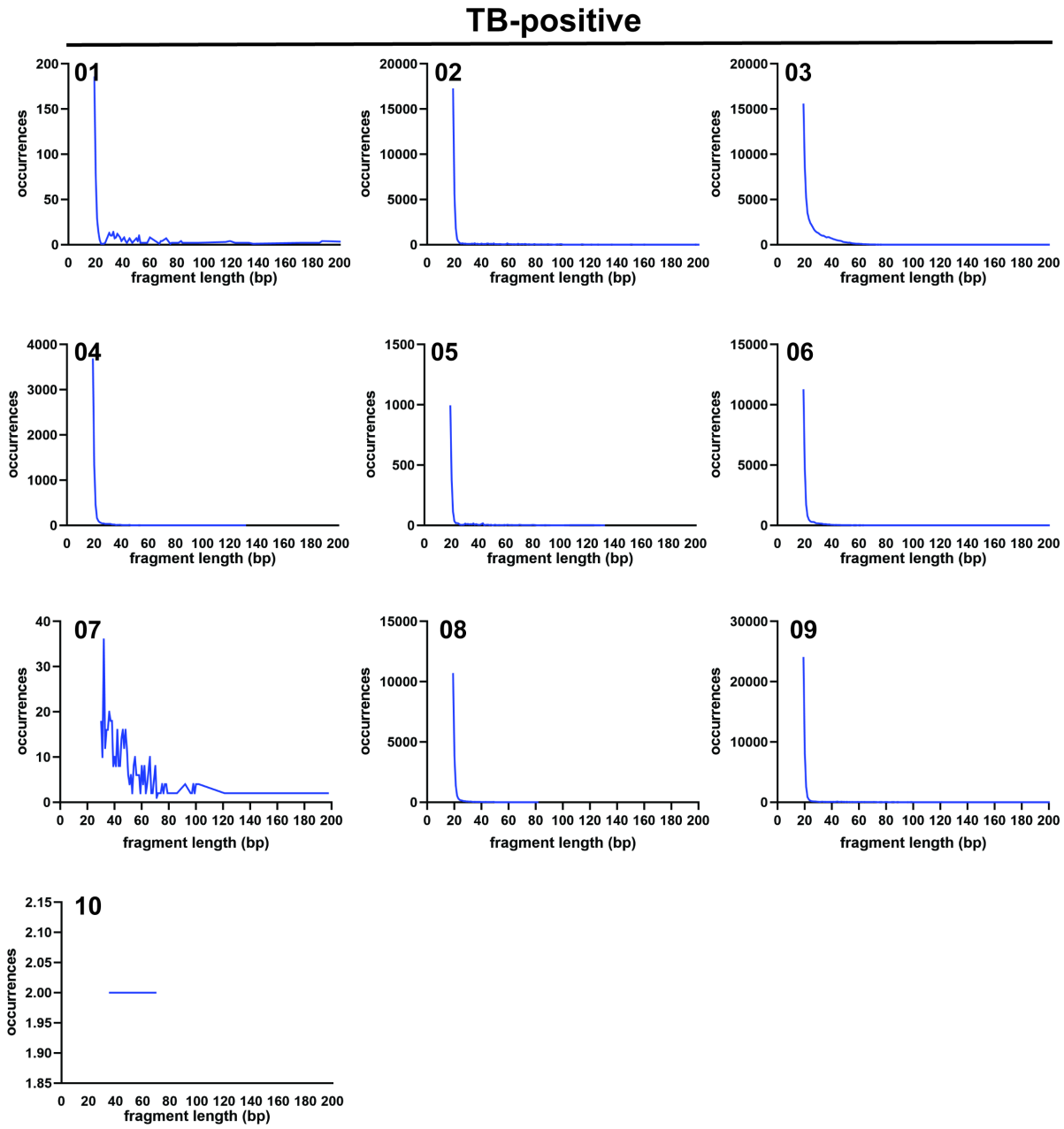


Figure 13.2: Fragment length distribution of cfDNA mapped to the TB genome in each sample. *Sample 10 had a low number of reads that mapped to the TB genome, but was confirmed to contain reads specific to TB by metagenomic classification. Neither TB-negative sample (Sample 11 or Sample 12, excluded here) contained reads specific to TB by metagenomic classification.*

14. REFERENCES

1. World Health Organization. 2020. Global tuberculosis report 2020.

2. World Health Organization. 2014. High-priority target product profiles for new tuberculosis diagnostics: report of a consensus meeting.

3. Botezatu I, Serdyuk O, Potapova G, Shelepov V, Alechina R, Molyaka Y, Ananév V, Bazin I, Garin A, Narimanov M, Knysh V, Melkonyan H, Umansky S, Lichtenstein A. 2000. Genetic analysis of DNA excreted in urine: a new approach for detecting specific genomic DNA sequences from cells dying in an organism. Clin Chem 46:1078–84.

4. Cannas A, Goletti D, Girardi E, Chiacchio T, Calvo L, Cuzzi G, Piacentini M, Melkonyan H, Umansky SR, Lauria FN, Ippolito G, Tomei LD. 2008. Mycobacterium tuberculosis DNA detection in soluble fraction of urine from pulmonary tuberculosis patients. Int J Tuberc Lung Dis 12:146–151.

5. Labugger I, Heyckendorf J, Dees S, Häussinger E, Herzmann C, Kohl TA, Richter E, Rivera-Milla E, Lange C. 2016. Detection of transrenal DNA for the diagnosis of pulmonary tuberculosis and treatment monitoring. Infection 45:1–8.

6. Patel K, Nagel M, Wesolowski M, Dees S, Rivera-Milla E, Geldmacher C, Dheda K, Hoelscher M, Labugger I. 2018. Evaluation of a Urine-Based Rapid Molecular Diagnostic Test with Potential to Be Used at Point-of-Care for Pulmonary Tuberculosis: Cape Town Cohort. J Mol Diagnostics 20:215–224.

7. Horne DJ, Kohli M, Zifodya JS, Schiller I, Dendukuri N, Tollefson D, Schumacher SG, Ochodo EA, Pai M, Steingart KR. 2019. Xpert MTB/RIF and Xpert MTB/RIF Ultra for pulmonary tuberculosis and rifampicin resistance in adults. Cochrane Database Syst Rev 2019:6.

8. Detjen AK, DiNardo AR, Leyden J, Steingart KR, Menzies D, Schiller I, Dendukuri N, Mandalakas AM. 2015. Xpert MTB/RIF assay for the diagnosis of pulmonary tuberculosis in children: a systematic review and meta-analysis. Lancet Respir Med 3:451–61.

9. Kohli M, Schiller I, Dendukuri N, Dheda K, Denkinger CM, Schumacher SG, Steingart KR. 2018. Xpert MTB/RIF assay for extrapulmonary tuberculosis and rifampicin resistance. Cochrane Database Syst Rev 2018:8.

10. Steingart KR, Ramsay A, Pai M. 2014. Optimizing sputum smear microscopy for the diagnosis of pulmonary tuberculosis. Expert Rev Anti Infect Ther.

11. Dorman SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, Hall SL, Chakravorty S, Cirillo DM, Tukvadze N, Bablishvili N, Stevens W, Scott L, Rodrigues C, Kazi MI, Joloba M, Nakiyingi L, Nicol MP, Ghebrekristos Y, Anyango I, Murithi W, Dietze R, Lyrio Peres R, Skrahina A, Auchynka V, Chopra KK, Hanif M, Liu X, Yuan X, Boehme CC, Ellner JJ, Denkinger CM, study team SE, Schumacher SG, Alland D, Nabeta P, Armstrong DT, King B, Hall SL, Chakravorty S, Cirillo DM, Tukvadze N, Bablishvili N, Stevens W, Scott L, Rodrigues C, Kazi MI, Joloba M, Nakiyingi L, Nicol MP, Ghebrekristos Y, Anyango I, Murithi W, Dietze R, Peres RL, Skrahina A, Auchynka V, Chopra KK, Hanif M, Liu X, Yuan X, Boehme CC, Ellner JJ, Denkinger CM, Manabe YC, Hom D, Aspindzelashvili R, David A, Surve U, Kamulegeya LH, Nabweyambo S, Surtie S, Hapeela N, Cain KP, Agaya J, McCarthy KD, Marques-Rodrigues P, Castellani LGS, Almeida PS, Aguiar PPL de, Solodovnikova V, Ruan X, Liang L, Zhang G, Zhu H, Xie Y. 2018. Xpert MTB/RIF Ultra for detection of Mycobacterium tuberculosis and rifampicin resistance: a prospective multicentre diagnostic accuracy study. Lancet Infect Dis 18:76–84.

12. WHO. 2018. HIV-Associated Tuberculosis Fact Sheet.

13. Zar HJ, Workman LJ, Prins M, Bateman LJ, Mbhele SP, Whitman CB, Denkinger CM, Nicol MP. 2019. Tuberculosis diagnosis in children using Xpert Ultra on different respiratory specimens. Am J Respir Crit Care Med 200:1531–1538.

14. Perez-Risco D, Rodriguez-Temporal D, Valledor-Sanchez I, Alcaidea F. 2018. Evaluation of the Xpert MTB/RIF Ultra Assay for Direct Detection of Mycobacterium tuberculosis Complex in Smear-Negative Extrapulmonary Samples. J Clin Microbiol 56.

15. Peter JG, Theron G, van Zyl-Smit R, Haripersad A, Mottay L, Kraus S, Binder A, Meldau R, Hardy A, Dheda K. 2012. Diagnostic accuracy of a urine lipoarabinomannan strip-test for TB detection in HIV-infected hospitalised patients. Eur Respir J 40:1211–20.

16. Thomas TA. 2017. Tuberculosis in Children. Pediatr Clin North Am. W.B. Saunders.

17. Peter JG, Theron G, Muchinga TE, Govender U, Dheda K. 2012. The diagnostic accuracy of urine-based Xpert MTB/RIF in HIV-infected hospitalized patients who are smear-negative or sputum scarce. PLoS One 7:1–8.

18. Shenai S, Amisano D, Ronacher K, Kriel M, Banada PP, Song T, Lee M, Joh JS, Winter J, Thayer R, Via LE, Kim S, Barry 3rd CE, Walzl G, Alland D, Barry CE, Walzl G, Alland D. 2013. Exploring Alternative Biomaterials for Diagnosis of Pulmonary Tuberculosis in HIV-Negative Patients by Use

of the GeneXpert MTB/RIF Assay. J Clin Microbiol2013/10/11. 51:4161–4166.

19. Lawn SD, Kerkhoff AD, Vogt M, Wood R. 2013. Europe PMC Funders Group High diagnostic yield of tuberculosis from screening urine samples from HIV-infected patients with advanced immunodeficiency using the Xpert MTB / RIF assay. J Acquir Immune Defic Syndr 60:289–294.

20. Bjerrum S, Schiller I, Dendukuri N, Kohli M, Nathavitharana RR, Zwerling AA, Denkinger CM, Steingart KR, Shah M. 2019. Lateral flow urine lipoarabinomannan assay for detecting active tuberculosis in people living with HIV. Cochrane Database Syst Rev 2019:10.

21. World Health Organization. 2019. Lateral flow urine lipoarabinomannan assay (LF-LAM) for the diagnosis of active tuberculosis in people living with HIV: Policy update (2019).

22. Broger T, Nicol MP, Sigal GB, Gotuzzo E, Zimmer AJ, Surtie S, Caceres-Nakiche T, Mantsoki A, Reipold EI, Székely R, Tsionsky M, van Heerden J, Plisova T, Chikamatsu K, Lowary TL, Pinter A, Mitarai S, Moreau E, Schumacher SG, Denkinger CM. 2020. Diagnostic accuracy of 3 urine lipoarabinomannan tuberculosis assays in HIV-negative outpatients. J Clin Invest 130:5756–5764.

23. Broger T, Sossen B, du Toit E, Kerkhoff AD, Schutz C, Ivanova Reipold E, Ward A, Barr DA, Macé A, Trollip A, Burton R, Ongarello S, Pinter A, Lowary TL, Boehme C, Nicol MP, Meintjes G, Denkinger CM. 2019. Novel lipoarabinomannan point-of-care tuberculosis test for people with HIV: a diagnostic accuracy study. Lancet Infect Dis 19:852–861.

24. Luabeya AK, Wood RC, Shenje J, Filander E, Ontong C, Mabwe S, Africa H, Nguyen FK, Olson A, Weigel KM, Jones-Engel L, Hatherill M, Cangelosi GA. 2019. Noninvasive detection of tuberculosis by oral swab analysis. J Clin Microbiol 57.

25. Nicol MP, Wood RC, Workman L, Prins M, Whitman C, Ghebrekristos Y, Mbhele S, Olson A, Jones-Engel LE, Zar HJ, Cangelosi GA. 2019. Microbiological diagnosis of pulmonary tuberculosis in children by oral swab polymerase chain reaction. Sci Rep 9.

26. Mesman AW, Rodriguez C, Ager E, Coit J, Trevisi L, Franke MF. 2019. Diagnostic accuracy of molecular detection of Mycobacterium tuberculosis in pediatric stool samples: A systematic review and meta-analysis. Tuberculosis 119:101878.

27. Flores JA, Calderón R, Mesman AW, Soto M, Coit J, Aliaga J, Mendoza M, Leon SR, Konda K, Mestanza FM, Mendoza CJ, Lecca L, Murray MB, Holmberg RC, Pollock NR, Franke MF. 2020. Detection of Mycobacterium tuberculosis DNA in buccal swab samples from children in Lima,

Peru. Pediatr Infect Dis J 39:E376–E380.

28. Mesman AW, Calderon R, Soto M, Coit J, Aliaga J, Mendoza M, Franke MF. 2019. Mycobacterium tuberculosis detection from oral swabs with Xpert MTB/RIF ULTRA: A pilot study. BMC Res Notes 12:349.

29. Mesman AW, Soto M, Coit J, Calderon R, Aliaga J, Pollock NR, Mendoza M, Mestanza FM, Mendoza CJ, Murray MB, Lecca L, Holmberg R, Franke MF. 2019. Detection of Mycobacterium tuberculosis in pediatric stool samples using TruTip technology. BMC Infect Dis 19:563.

30. Gaur M, Singh A, Sharma V, Tandon G, Bothra A, Vasudeva A, Kedia S, Khanna A, Khanna V, Lohiya S, Varma-Basil M, Chaudhry A, Misra R, Singh Y. 2020. Diagnostic performance of non-invasive, stool-based molecular assays in patients with paucibacillary tuberculosis. Sci Rep 10:7102.

31. MANDEL P, METAIS P. 1948. Les acides nucléiques du plasma sanguin chez l'homme. C R Seances Soc Biol Fil 142:241–3.

32. Leon SA, Shapiro B, Sklaroff DM, Yaros MJ. 1977. Free DNA in the serum of cancer patients and the effect of therapy. Cancer Res 37:646–50.

33. Vasioukhin V, Anker P, Maurice P, Lyautey J, Lederrey C, Stroun M. 1994. Point mutations of the N-ras gene in the blood plasma DNA of patients with myelodysplastic syndrome or acute myelogenous leukaemia. Br J Haematol 86:774–779.

34. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. 1994. Soluble normal and mutated DNA sequences from single-copy genes in human blood. Cancer Epidemiol Biomarkers Prev 3:67–71.

35. Lo YMD, Corbetta N, Chamberlain PF, Rai V, Sargent IL, Redman CW, Wainscoat JS. 1997. Presence of fetal DNA in maternal plasma and serum. Lancet 350:485–487.

36. Lo YM, Tein MS, Pang CC, Yeung CK, Tong KL, Hjelm NM. 1998. Presence of donor-specific DNA in plasma of kidney and liver-transplant recipients. Lancet (London, England) 351:1329–30.

37. Aucamp J, Bronkhorst AJ, Badenhorst CPS, Pretorius PJ. 2018. The diverse origins of circulating cell-free DNA in the human body: a critical re-evaluation of the literature. Biol Rev 93:1649–1683.

38. Reckamp KL, Melnikova VO, Karlovich C, Sequist L V., Camidge DR, Wakelee H, Perol M, Oxnard GR, Kosco K, Croucher P, Samuelsz E, Vibat CR, Guerrero S, Geis J, Berz D, Mann E, Matheny S,

Rolfe L, Raponi M, Erlander MG, Gadgeel S. 2016. A highly sensitive and quantitative test platform for detection of NSCLC EGFR mutations in urine and plasma. J Thorac Oncol 11:1690–1700.

39. Fujii T, Barzi A, Sartore-Bianchi A, Cassingena A, Fujii T, Barzi A, Sartore-Bianchi A, Cassingena A, Siravegna G, Karp DD, Piha-Paul SA, Subbiah V, Tsimberidou AM, Huang HJ, Veronese S, Di Nicolantonio F, Pingle S, Vibat CRT, Hancock S, Berz D, Melnikova VO, Erlander MG, Luthra R, Kopetz ES, Meric-Bernstam F, Siena S, Lenz H-J, Bardelli A, Janku F. 2017. Mutation-Enrichment Next-Generation Sequencing for Quantitative Detection of KRAS Mutations in Urine Cell-Free DNA from Patients with Advanced Cancers. Clin Cancer Res 23:3657–3666.

40. Chakrabarti AK, Caruso L, Ding M, Shen C, Buchanan W, Gupta P, Rinaldo CR, Chen Y. 2009. Detection of HIV-1 RNA/DNA and CD4 mRNA in feces and urine from chronic HIV-1 infected subjects with and without anti-retroviral therapy. AIDS Res Ther2009/10/06. 6:20.

41. Weerakoon KG, McManus DP. 2016. Cell-Free DNA as a Diagnostic Tool for Human Parasitic Infections. Trends Parasitol 32:378–391.

42. Leticia Fernández-Carballo B, Broger T, Wyss R, Banaei N. 2019. Towards the development of a cfDNA-based in-vitro diagnostic test for infectious diseases: A review of evidence for tuberculosis. J Clin Microbiol 57:e01234-18.

43. Zhang J, Tong K-L, Li PKT, Chan AYW, Yeung C-K, Pang CCP, Wong TYH, Lee K-C, Dennis Lo YM. 1999. Presence of Donor-and Recipient-derived DNA in Cell-free Urine Samples of Renal Transplantation Recipients: Urinary DNA Chimerism. Clin Chem 45:1741–1746.

44. Gielis EM, Ledeganck KJ, De Winter BY, Del Favero J, Bosmans J-LL, Claas FHJ, Abramowicz D, Eikmans M. 2015. Cell-Free DNA: An Upcoming Biomarker in Transplantation. Am J Transplant 15:2541–2551.

45. Tsui NBY, Jiang P, Chow KCK, Su X, Leung TY, Sun H, Chan KCA, Chiu RWK, Lo YMD. 2012. High resolution size analysis of fetal DNA in the urine of pregnant women by paired-end massively parallel sequencing. PLoS One 7:e48319.

46. Yu SCY, Lee SWY, Jiang P, Leung TY, Chan KCA, Chiu RWK, Lo YMD, Redman CW. 2013. High-resolution profiling of fetal DNA clearance from maternal plasma by massively parallel sequencing. Clin Chem 59:1228–37.

47. Su Y-H, Wang M, Brenner DE, Ng A, Melkonyan H, Umansky S, Syngal S, Block TM. 2004. Human urine contains small, 150 to 250 nucleotide-sized, soluble DNA derived from the circulation and may be useful in the detection of colorectal cancer. J Mol Diagn 6:101–7.

48. Streleckiene G, Reid HM, Arnold N, Bauerschlag D, Forster M. 2018. Quantifying cell free DNA in urine: comparison between commercial kits, impact of gender and inter-individual variation. Biotechniques 64:225–230.

49. Burnham P, Kim MS, Agbor-Enoh S, Luikart H, Valantine HA, Khush KK, De Vlaminck I. 2016. Single-stranded DNA library preparation uncovers the origin and diversity of ultrashort cell-free DNA in plasma. Sci Rep 6:27859.

50. Underhill HR, Kitzman JO, Hellwig S, Welker NC, Daza R, Baker DN, Gligorich KM, Rostomily RC, Bronner MP, Shendure J. 2016. Fragment Length of Circulating Tumor DNA. PLoS Genet 12:1–24.

51. Snyder MW, Kircher M, Hill AJ, Daza RM, Shendure J. 2016. Cell-free DNA Comprises an in Vivo Nucleosome Footprint that Informs Its Tissues-Of-Origin. Cell 164:57–68.

52. Green C, Huggett JF, Talbot E, Mwaba P, Reither K, Zumla AI. 2009. Rapid diagnosis of tuberculosis through the detection of mycobacterial DNA in urine by nucleic acid amplification methods. Lancet Infect Dis 9:505–511.

53. Yao W, Mei C, Nan X, Hui L. 2016. Evaluation and comparison of in vitro degradation kinetics of DNA in serum, urine and saliva: A qualitative study. Gene 590:142–148.

54. Cheng THTT, Jiang P, Tam JCWW, Sun X, Lee W-S, Yu SCYY, Teoh JYCC, Chiu PKFF, Ng C-F, Chow K-M, Szeto C-C, Chan KCACA, Chiu RWKK, Dennis Lo YM, Lo YMD. 2017. Genomewide bisulfite sequencing reveals the origin and time-dependent fragmentation of urinary cfDNA. Clin Biochem 50:496–501.

55. Burnham P, Dadhania D, Heyang M, Chen F, Westblade LF, Suthanthiran M, Lee JR, De Vlaminck I. 2018. Urinary cell-free DNA is a versatile analyte for monitoring infections of the urinary tract. Nat Commun 9:2412.

56. Markus H, Zhao J, Contente-Cuomo T, Stephens MD, Raupach E, Odenheimer-Bergman A, Connor S, McDonald BR, Moore B, Hutchins E, McGilvrey M, de la Maza MC, Van Keuren-Jensen K, Pirrotte P, Goel A, Becerra C, Von Hoff DD, Celinski SA, Hingorani P, Murtaza M. 2021. Analysis of recurrently protected genomic regions in cell-free DNA found in urine. Sci Transl Med

13:eaaz3088.

57. Crisafulli G, Mussolin B, Cassingena A, Montone M, Bartolini A, Barault L, Martinetti A, Morano F, Pietrantonio F, Sartore-Bianchi A, Siena S, Di Nicolantonio F, Marsoni S, Bardelli A, Siravegna G. 2019. Whole exome sequencing analysis of urine trans-renal tumour DNA in metastatic colorectal cancer patients. ESMO Open 4:e000572.

58. Sanchez C, Snyder MW, Tanos R, Shendure J, Thierry AR. 2018. New insights into structural features and optimal detection of circulating tumor DNA determined by single-strand DNA analysis. npj Genomic Med 3:31.

59. Bordelon H, Ricks KM, Pask ME, Russ PK, Solinas F, Baglia ML, Short PA, Nel A, Blackburn J, Dheda K, Zamudio C, Cáceres T, Wright DW, Haselton FR, Pettit AC. 2017. Design and use of mouse control DNA for DNA biomarker extraction and PCR detection from urine: Application for transrenal Mycobacterium tuberculosis DNA detection. J Microbiol Methods 136:65–70.

60. Fortún J, Martín-Dá P, G ´ Omez-Mampaso E, Gonzá Lez-García A, Barbolla I, G ´ Omez-García I, Wikman P, Ortíz J, Navas E, Cuartero C, Gijón D, Gijón G, Moreno S. 2014. Extra-pulmonary tuberculosis: differential aspects and role of 16S-rRNA in urine. Int J Tuberc Lung Dis 18:478–485.

61. Click ES, Murithi W, Ouma GS, McCarthy K, Willby M, Musau S, Alexander H, Pevzner E, Posey J, Cain KP. 2018. Detection of Apparent Cell-free M. tuberculosis DNA from Plasma. Sci Rep 8:645.

62. Ushio R, Yamamoto M, Nakashima K, Watanabe H, Nagai K, Shibata Y, Tashiro K, Tsukahara T, Nagakura H, Horita N, Sato T, Shinkai M, Kudo M, Ueda A, Kaneko T. 2016. Digital PCR assay detection of circulating Mycobacterium tuberculosis DNA in pulmonary tuberculosis patient plasma. Tuberculosis 99:47–53.

63. Sechi LA, Pinna MP, Sanna A, Pirina P, Ginesu F, Saba F, Aceti A, Turrini F, Zanetti S, Fadda G. 1997. Detection of Mycobacterium tuberculosis by PCR analysis of urine and other clinical samples from AIDS and non-HIV-infected patients. Mol Cell Probes 11:281–5.

64. Aceti A, Zanetti S, Mura MS, Sechi LA, Turrini F, Saba F, Babudieri S, Mannu F, Fadda G. 1999. Identification of HIV patients with active pulmonary tuberculosis using urine based polymerase chain reaction assay. Thorax 54:145–6.

65. Kafwabulula M, Ahmed K, Nagatake T, Gotoh J, Mitarai S, Oizumi K ZA. 2002. Evaluation of PCR-based methods for the diagnosis of tuberculosis by identifi cation of mycobacterial DNA in urine

samples. Int J Tuberc Lung Dis 6:732–7.

66. Torrea G, Van De Perre P, Ouedraogo M, Zougba A, Sawadogo A, Dingtoumda BBB, Diallo B, Defer MC, Sombié I, Zanetti S, Sechi LA. 2005. PCR-based detection of the Mycobacterium tuberculosis complex in urine of HIV-infected and uninfected pulmonary and extrapulmonary tuberculosis patients in Burkina Faso. J Med Microbiol 54:39–44.

67. Rebollo MJ, San Juan Garrido R, Folgueira D, Palenque E, Díaz-Pedroche C, Lumbreras C, Aguado JM. 2006. Blood and urine samples as useful sources for the direct detection of tuberculosis by polymerase chain reaction. Diagn Microbiol Infect Dis 56:141–146.

68. Gopinath K, Singh S. 2009. Urine as an adjunct specimen for the diagnosis of active pulmonary tuberculosis. Int J Infect Dis 13:374–379.

69. da Cruz HLA, de Albuquerque Montenegro R, de Araújo Lima JF, da Rocha Poroca D, da Costa Lima JF, Maria Lapa Montenegro L, Crovella S, Charifker Schindler H, Montenegro LML, Crovella S, Schindler HC. 2011. Evaluation of a nested-PCR for mycobacterium tuberculosis detection in blood and urine samples. Braz J Microbiol 42:321–329.

70. Amin I, Idrees M, Awan Z, Shahid M, Afzal S, Hussain A. 2011. PCR could be a method of choice for identification of both pulmonary and extra-pulmonary tuberculosis. BMC Res Notes 4:332.

71. Lima JF da C, Guedes G de MR, Lima JF de A, Lira LA de S, Santos FCF, Arruda ME de, Montenegro LML, Schindler HC. 2015. Single-tube nested PCR assay with in-house DNA extraction for Mycobacterium tuberculosis detection in blood and urine. Rev Soc Bras Med Trop 48:731–8.

72. Kafwabulula M, Ahmed K, Nagatake T, Gotoh J, Mitarai S, Oizumi K, Zumla A. 2002. Evaluation of PCR-based methods for the diagnosis of tuberculosis by identification of mycobacterial DNA in urine samples. Int J Tuberc Lung Dis 6:732–7.

73. Boom R, Sol C, MM S, Jansen C, Wertheim-van Dillen P, van der Noordaa J. 1990. Rapid and Simple Method for Purification of Nucleic Acids. J Clin Microbiol 28:495–503.

74. Melzak KA, Sherwood CS, Turner RFB, Haynes CA. 1996. Driving Forces for DNA Adsorption to Silica in Perchlorate Solutions. J Colloid Interface Sci 181:635–644.

75. Katevatis C, Fan A, Klapperich CM. 2017. Low concentration DNA extraction and recovery using a silica solid phase. PLoS One 12:e0176848.

76. Vandeventer PE, Lin JS, Zwang TJ, Nadim A, Johal MS, Niemz A. 2012. Multiphasic DNA

Adsorption to Silica Surfaces under Varying Buffer, pH, and Ionic Strength Conditions. J Phys Chem B 116:5661–5670.

77. Bordelon H, Russ PK, Wright DW, Haselton FR. 2013. A magnetic bead-based method for concentrating DNA from human urine for downstream detection. PLoS One 8:e68369.

78. Su Y-H, Wang M, Block TM, Landt O, Botezatu I, Serdyuk O, Lichtenstein A, Melkonyan H, Tomei LD, Umansky S. 2004. Transrenal DNA as a diagnostic tool: important technical notes. Ann N Y Acad Sci 1022:81–9.

79. Shekhtman EM, Anne K, Melkonyan HS, Robbins DJ, Warsof SL, Umansky SR. 2009. Optimization of transrenal DNA analysis: detection of fetal DNA in maternal urine. Clin Chem 55:723–9.

80. Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. 2008. Quantitative Analysis of the Transrenal Excretion of Circulating EBV DNA in Nasopharyngeal Carcinoma Patients. Clin Cancer Res 14:4809–4813.

81. Melkonyan HS, Feaver WJ, Meyer E, Scheinker V, Shekhtman EM, Xin ZH, Umansky SR. 2008. Transrenal Nucleic Acids: From Proof of Principle to Clinical Tests. Ann N Y Acad Sci 1137:73–81.

82. Marangu D, Devine B, John-Stewart G. 2015. Diagnostic accuracy of nucleic acid amplification tests in urine for pulmonary tuberculosis: A meta-analysis. Int J Tuberc Lung Dis 19:1339–47.

83. Petrucci R, Lombardi G, Corsini I, Visciotti F, Pirodda A, Cazzato S, Landini MP, Dal Monte P. 2015. Use of transrenal DNA for the diagnosis of extrapulmonary tuberculosis in children: a case of tubercular otitis media. J Clin Microbiol 53:336–8.

84. Yamamoto M, Ushio R, Watanabe H, Tachibana T, Tanaka M, Yokose T, Tsukiji J, Nakajima H, Kaneko T. 2018. Detection of Mycobacterium tuberculosis-derived DNA in circulating cell-free DNA from a patient with disseminated infection using digital PCR. Int J Infect Dis 66:80–82.

85. Yang J, Han X, Liu A, Bai X, Xu C, Bao F, Feng S, Tao L, Ma M, Peng Y. 2017. Use of Digital Droplet PCR to Detect Mycobacterium tuberculosis DNA in Whole Blood-Derived DNA Samples from Patients with Pulmonary and Extrapulmonary Tuberculosis. Front Cell Infect Microbiol 7:369.

86. da Costa-Lima JF, Pimentel LMLM, Santos FCF, Salazar MP, Duarte RS, Mello FC de Q, Schindler HC. 2020. Rapid detection of mycobacterium tuberculosis in children using blood and urine specimens. Rev Soc Bras Med Trop 53:1–9.

87. Cannas A, Kalunga G, Green C, Calvo L, Katemangwe P, Reither K, Perkins MD, Maboko L,

Hoelscher M, Talbot EA, Mwaba P, Zumla AI, Girardi E, Huggett JF, trDNA Consortium for the TB. 2009. Implications of Storing Urinary DNA from Different Populations for Molecular Analyses. PLoS One 4:e6985.

88. Murugesan K, Hogan CA, Palmer Z, Reeve B, Theron G, Andama A, Somoskovi A, Steadman A, Madan D, Andrews J, Croda J, Sahoo MK, Cattamanchi A, Pinsky BA, Banaei N. 2019. Investigation of preanalytical variables impacting pathogen cell-free DNA in blood and urine. J Clin Microbiol 57:e00782-19.

89. Milde A, Haas-Rochholz H, Kaatsch H-J. Improved DNA typing of human urine by adding EDTA.

90. Bosschieter J, Bach S, Bijnsdorp I V., Segerink LI, Rurup WF, van Splunter AP, Bahce I, Novianti PW, Kazemier G, van Moorselaar RJAA, Steenbergen RDMM, Nieuwenhuijzen JA. 2018. A protocol for urine collection and storage prior to DNA methylation analysis. PLoS One 13:e0200906.

91. Lee EY, Lee E-J, Yoon H, Lee DH, Kim KH. 2020. Comparison of Four Commercial Kits for Isolation of Urinary Cell-Free DNA and Sample Storage Conditions. Diagnostics 10:234.

92. Augustus E, Van Casteren K, Sorber L, van Dam P, Roeyen G, Peeters M, Vorsters A, Wouters A, Raskin J, Rolfo C, Zwaenepoel K, Pauwels P. 2020. The art of obtaining a high yield of cell-free DNA from urine. PLoS One 15:e0231058.

93. El Bali L, Diman A, Bernard A, Roosens NHC, De Keersmaecker SCJ. 2014. Comparative study of seven commercial kits for human DNA extraction from urine samples suitable for DNA biomarker-based public health studies. J Biomol Tech 25:96–110.

94. Chen J, Johnson R, Griffiths M. 1998. Detection of verotoxigenic Escherichia coli by magnetic capture-hybridization PCR. Appl Environ Microbiol 64:147–52.

95. Chen J, Griffiths MW. 2008. Detection of Salmonella and simultaneous detection of Salmonella and Shiga-like toxin-producing Escherichia coli using the magnetic capture hybridization polymerase chain reaction. Lett Appl Microbiol 32:7–11.

96. Johnson KL, Zheng D, Kaewnum S, Reid CL, Burr T. 2013. Development of a Magnetic Capture Hybridization Real-Time PCR Assay for Detection of Tumorigenic Agrobacterium vitis in Grapevines 103:633.

97. Bach H-J, Hartmann A, Trevors JT, Munch JC. 1999. Magnetic capture–hybridization method for

purification and probing of mRNA for neutral protease of Bacillus cereus. J Microbiol Methods 37:187–192.

98. Millar DS, Withey SJ, Tizard MLV, Ford JG, Hermontaylor J. 1995. Solid-Phase Hybridization Capture of Low-Abundance Target DNA Sequences: Application to the Polymerase Chain Reaction Detection of Mycobacterium paratuberculosis and Mycobacterium avium subsp. silvaticum. Anal Biochem 226:325–330.

99. Marsh IB, Whittington RJ. 2001. Progress towards a rapid polymerase chain reaction diagnostic test for the identification of Mycobacterium avium subsp. paratuberculosis in faeces. Mol Cell Probes 15:105–118.

100. Maher N, Dillon HK, Vermund SH, Unnasch TR. 2001. Magnetic bead capture eliminates PCR inhibitors in samples collected from the airborne environment, permitting detection of Pneumocystis carinii DNA. Appl Environ Microbiol 67:449–52.

101. Langrell SRH, Barbara DJ. 2001. Magnetic capture hybridisation for improved PCR detection ofNectria galligena from lignified apple extracts. Plant Mol Biol Report 19:5–11.

102. Maibach RC, Dutly F, Altwegg M. 2002. Detection of Tropheryma whipplei DNA in feces by PCR using a target capture method. J Clin Microbiol 40:2466–71.

103. Ha Y, Fessehaie A, Ling KS, Wechter WP, Keinath AP, Walcott RR. 2009. Simultaneous Detection of Acidovorax avenae subsp. citrulli and Didymella bryoniae in Cucurbit Seedlots Using Magnetic Capture Hybridization and Real-Time Polymerase Chain Reaction 99:666–678.

104. Bai Y, Cui Y, Suo Y, Shi C, Wang D, Shi X. 2019. A Rapid Method for Detection of Salmonella in Milk Based on Extraction of mRNA Using Magnetic Capture Probes and RT-qPCR. Front Microbiol 10:770.

105. Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust EM, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C. 2009. Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. Nat Biotechnol 27:182–189.

106. Albretsen C, Kalland K-H, Haukanes B-I, Håvarstein L-S, Kleppe K. 1990. Applications of magnetic beads with covalently attached oligonucleotides in hybridization: Isolation and detection of specific measles virus mRNA from a crude cell lysate. Anal Biochem 189:40–50.

107. Beaulieux F, See DM, Leparc-Goffart I, Aymard M, Lina B. 1997. Use of magnetic beads versus guanidium thiocyanate-phenol-chloroform rna extraction followed by polymerase chain reaction for the rapid, sensitive detection of enterovirus rna. Res Virol 148:11–15.

108. Yeung SW, Hsing I-M. 2006. Manipulation and extraction of genomic DNA from cell lysate by functionalized magnetic particles for lab on a chip applications. Biosens Bioelectron 21:989–997.

109. Rohrman B, Richards-Kortum R. 2015. Inhibition of Recombinase Polymerase Amplification by Background DNA: A Lateral Flow-Based Method for Enriching Target DNA. Anal Chem 87:1963–1967.

110. Muir P, Nicholson F, Jhetam M, Neogi S, Banatvala JE. 1993. Rapid diagnosis of enterovirus infection by magnetic bead extraction and polymerase chain reaction detection of enterovirus RNA in clinical specimens. J Clin Microbiol 31:31–8.

111. Heermann K-H, Hagos Y, Thomssen R. 1994. Liquid-phase hybridization and capture of hepatitis B virus DNA with magnetic beads and fluorescence detection of PCR product. J Virol Methods 50:43–57.

112. Jacobsen CS. 1995. Microscale detection of specific bacterial DNA in soil with a magnetic capture-hybridization and PCR amplification assay. Appl Environ Microbiol 61:3347–52.

113. Mangiapan G, Vokurka M, Schouls L, Cadranel J, Lecossier D, Van Embden J, Hance AJ. 1996. Sequence Capture-PCR Improves Detection of Mycobacterial DNA in Clinical Specimens. J Clin Microbiol 34:1209–1215.

114. Brugière O, Vokurka M, Lecossier D, Mangiapan G, Amrane A, Milleron B, Mayaud C, Cadranel J, Hance AJ. 1997. Diagnosis of smear-negative pulmonary tuberculosis using sequence capture polymerase chain reaction. Am J Respir Crit Care Med 155:1478–1481.

115. Arnal C, Ferré-Aubineau V, Besse B, Mignotte B, Schwartzbrod L, Billaudel S. 1999. Comparison of seven RNA extraction methods on stool and shellfish samples prior to hepatitis A virus amplification. J Virol Methods 77:17–26.

116. Amagliani G, Omiccioli E, Campo A, Bruce IJ, Brandi G, Magnani M. 2006. Development of a magnetic capture hybridization-PCR assay for Listeria monocytogenes direct detection in milk samples. J Appl Microbiol 100:375–383.

117. Thompson DE, Rajal VB, De Batz S, Wuertz S. 2006. Detection of Salmonella spp. in water using

magnetic capture hybridization combined with PCR or real-time PCR. J Water Health 4:67–75.

118. Parham NJ, Ois F, Picard J, Peytavi R, Gagnon M, Seyrig G, Gagné P-A, Boissinot M, Bergeron MG. 2007. Specific Magnetic Bead–Based Capture of Genomic DNA from Clinical Samples: Application to the Detection of Group B Streptococci in Vaginal/Anal Swabs. Clin Chem 53:1570–1576.

119. Vansnick E, De Rijk P, Vercammen F, Rigouts L, Portaels F, Geysen D. 2007. A DNA sequence capture extraction method for detection of subspecies in feces and tissue samples. Vet Microbiol 122:166.

120. Peeters S, Stakenborg T, Colle F, Liu CX, Lagae L, Ranst M. 2012. Specific magnetic isolation for direct detection of HPV16. Eur J Clin Microbiol Infect Dis 31:539–546.

121. Rodriguez D, Longo A V., Zamudio KR. 2012. Magnetic capture hybridization of Batrachochytrium dendrobatidis genomic DNA. J Microbiol Methods 90:156–159.

122. Adams NM, Bordelon H, Wang K-KA, Albert LE, Wright DW, Haselton FR. 2015. Comparison of Three Magnetic Bead Surface Functionalities for RNA Extraction and Detection. ACS Appl Mater Interfaces 7:6062–6069.

123. Guo Y, Sheng S, Nie B, Tu Z. 2015. Development of magnetic capture hybridization and quantitative polymerase chain reaction for hepatitis B virus covalently closed circular DNA. Hepat Mon 15:e23729.

124. Reed JL, Walker ZJ, Basu D, Allen V, Nicol MP, Kelso DM, McFall SM. 2016. Highly sensitive sequence specific qPCR detection of Mycobacterium tuberculosis complex in respiratory specimens. Tuberculosis 101:114–124.

125. Reed JL, Basu D, Butzler MA, McFall SM. 2017. XtracTB Assay, a Mycobacterium tuberculosis molecular screening test with sensitivity approaching culture. Sci Rep 7:3653.

126. Lu T, Li J. 2017. Clinical applications of urinary cell-free DNA in cancer: current insights and promising future. Am J Cancer Res 7:2318–2332.

127. Oreskovic A, Brault ND, Panpradist N, Lai JJ, Lutz BR. 2019. Analytical Comparison of Methods for Extraction of Short Cell-Free DNA from Urine. J Mol Diagnostics 21:1067–1078.

128. Thierry D, Cave MD, Eisenach KD, Crawford JT, Bates JH, Gicquel B, Guesdon JL. 1990. *IS6110* , an IS-like element of *Mycobacterium tuberculosis* complex. Nucleic Acids Res 18:188–188.

129. del Carmen Menéndez M, Samper S, Otal I, García MJ. 2012. IS6110 the Double-Edged Passenger, p. 59–88. *In* Cardona, P-J (ed.), Understanding Tuberculosis – Deciphering the Secret Life of the Bacilli. InTech Open.

130. Hellyer TJ, Desjardin LE, Assaf MK, Bates JH, Cave MD, Eisenach KD. 1996. Specificity of IS6110-based amplification assays for Mycobacterium tuberculosis complex. J Clin Microbiol 34:2843–2846.

131. Alguacil J, Pfeiffer RM, Moore LE, del Fresno MR, Medina-Lopez R, Kogevinas M, Vermeulen R, Dosemeci M, Silverman DT, Rothman N, García-Closas M. 2007. Measurement of urine pH for epidemiological studies on bladder cancer. Eur J Epidemiol 22:91–98.

132. Wang C-Y, Cogswell ME, Loria CM, Chen T-C, Pfeiffer CM, Swanson CA, Caldwell KL, Perrine CG, Carriquiry AL, Liu K, Sempos CT, Gillespie CD, Burt VL. 2013. Urinary Excretion of Sodium, Potassium, and Chloride, but Not Iodine, Varies by Timing of Collection in a 24-Hour Calibration Study. J Nutr 143:1276–1282.

133. Radomski N, Kreitmann L, McIntosh F, Behr MA. 2013. The Critical Role of DNA Extraction for Detection of Mycobacteria in Tissues. PLoS One 8:e78749.

134. Sekar MMA, Bloch W, St John PM. 2005. Comparative study of sequence-dependent hybridization kinetics in solution and on microspheres. Nucleic Acids Res 33:366–75.

135. Henry MR, Wilkins Stevens P, Sun J, Kelso DM. 1999. Real-time measurements of DNA hybridization on microparticles with fluorescence resonance energy transfer. Anal Biochem 276:204–214.

136. Peterson AW, Heaton RJ, Georgiadis RM. 2001. The effect of surface probe density on DNA hybridization. Nucleic Acids Res 29:5163–8.

137. Ravan H, Kashanian S, Sanadgol N, Badoei-Dalfard A, Karami Z. 2014. Strategies for optimizing DNA hybridization on surfaces. Anal Biochem. Academic Press.

138. Steel AB, Herne TM, Tarlov MJ, H-j A, Biomol Struct AJ. 1997. Electrochemical Quantitation of DNA Immobilized on Gold. Proc Natl Acad Sci USA 69:47.

139. Huang E, Satjapipat M, Han S, Zhou F. 2001. Surface structure and coverage of an oligonucleotide probe tethered onto a gold substrate and its hybridization efficiency for a polynucleotide target. Langmuir 17:1215–1224.

140. Shchepinov MS, Case-Green SC, Southern EM. 1997. Steric factors influencing hybridisation of nucleic acids to oligonucleotide arrays. Nucleic Acids Res 25:1155.

141. Vainrub A, Pettitt BM. 2011. Accurate prediction of binding thermodynamics for DNA on surfaces. J Phys Chem B 115:13300–3.

142. González M, Argaraña CE, Fidelio GD. 1999. Extremely high thermal stability of streptavidin and avidin upon biotin binding. Biomol Eng 16:67–72.

143. Yuce M, Kurt H, Budak H. 2014. Characterization of a dual biotin tag for improved single stranded DNA production. Anal Methods 6:548–557.

144. Dressman D, Yan H, Traverso G, Kinzler KW, Vogelstein B. 2003. Transforming single DNA molecules into fluorescent magnetic particles for detection and enumeration of genetic variations. PNAS 100:8817–8822.

145. Chivers CE, Crozat E, Chu C, Moy VT, Sherratt DJ, Howarth M. 2010. A streptavidin variant with slower biotin dissociation and increased mechanostability. Nat Methods 7:391–393.

146. Samarasinghe S, Meah F, Singh V, Basit A, Emanuele N, Emanuele MA, Mazhari A, Holmes EW. 2017. Biotin interference with routine clinical immunoassays: Understand the causes and mitigate the risks. Endocr Pract. American Association of Clinical Endocrinologists.

147. Owczarzy R, You Y, Groth CL, Tataurov A V. 2011. Stability and mismatch discrimination of locked nucleic acid-DNA duplexes. Biochemistry 50:9352–9367.

148. You Y, Moreira BG, Behlke MA, Owczarzy R. 2006. Design of LNA probes that improve mismatch discrimination. Nucleic Acids Res 34.

149. Levin JD, Fiala D, Samala MF, Kahn JD, Peterson RJ. 2006. Position-dependent effects of locked nucleic acid (LNA) on DNA sequencing and PCR primers. Nucleic Acids Res 34.

150. Latorra D, Arar K, Hurley JM. 2003. Design considerations and effects of LNA in PCR primers. Mol Cell Probes 17:253–259.

151. Van Ness J, Chen L. 1991. The use of oligodeoxynucleotide probes in chaotrope-based hybridization solutions. Nucleic Acids Res 19:5143–51.

152. Trigg RM, Martinson LJ, Parpart-Li S, Shaw JA. 2018. Factors that influence quality and yield of circulating-free DNA: A systematic review of the methodology literature. Heliyon 4:e00699.

153. Lee DH, Yoon H, Park S, Kim JS, Ahn Y-H, Kwon K, Lee D, Kim KH. 2018. Urinary Exosomal and cell-free DNA Detects Somatic Mutation and Copy Number Alteration in Urothelial Carcinoma of Bladder. Sci Rep 8:14707.

154. Kishore R, Reef Hardy W, Anderson VJ, Sanchez NA, Buoncristiani MR. 2006. Optimization of DNA Extraction from Low-Yield and Degraded Samples Using the BioRobot ® EZ1 and BioRobot ® M48. J Forensic Sci 51:1055–1061.

155. Huggett JF, Novak T, Garson JA, Green C, Morris-Jones SD, Miller RF, Zumla A. 2008. Differential susceptibility of PCR reactions to inhibitors: an important and unrecognised phenomenon. BMC Res Notes 1:70.

156. Page K, Guttery DS, Zahra N, Primrose L, Elshaw SR, Pringle JH, Blighe K, Marchese SD, Hills A, Woodley L, Stebbing J, Coombes RC, Shaw JA. 2013. Influence of Plasma Processing on Recovery and Analysis of Circulating Nucleic Acids. PLoS One 8:e77963.

157. Devonshire AS, Whale AS, Gutteridge A, Jones G, Cowen S, Foy CA, Huggett JF. 2014. Towards standardisation of cell-free DNA measurement in plasma: controls for extraction efficiency, fragment size bias and quantification. Anal Bioanal Chem 406:6499–6512.

158. Sorber L, Zwaenepoel K, Deschoolmeester V, Roeyen G, Lardon F, Rolfo C, Pauwels P. 2017. A Comparison of Cell-Free DNA Isolation Kits: Isolation and Quantification of Cell-Free DNA in Plasma. J Mol Diagnostics 19:162–168.

159. Diefenbach RJ, Lee JH, Kefford RF, Rizos H. 2018. Evaluation of commercial kits for purification of circulating free DNA. Cancer Genet 228–229:21–27.

160. Shi B, Shin YK, Hassanali AA, Singer SJ. 2015. DNA Binding to the Silica Surface. J Phys Chem B 119:11030–11040.

161. Beld M, Sol C, Goudsmit J, Boom R. 1996. Fractionation of nucleic acids into single-stranded and double-stranded forms. Nucleic Acids Res 24:2618–2619.

162. Oreskovic A, Lutz B. 2021. Ultrasensitive hybridization capture: reliable detection of <1 copy/mL short cell-free DNA from large-volume urine samples. PLoS One 16:e0247851.

163. Oreskovic A, Panpradist N, Marangu D, Ngwane MW, Magcaba ZP, Ngcobo S, Ngcobo Z, Horne DJ, Wilson DPK, Shapiro AE, Drain PK, Lutz BR. 2021. Diagnosing pulmonary tuberculosis using sequence-specific purification of urine cell-free DNA. J Clin Microbiol 59:e00074-21.

164. Lyu L, Li Z, Pan L, Jia H, Sun Q, Liu Q, Zhang Z. 2020. Evaluation of digital PCR assay in detection of M.tuberculosis IS6110 and IS1081 in tuberculosis patients plasma. BMC Infect Dis 20:657.

165. Minion J, Leung E, Talbot E, Dheda K, Pai M, Menzies D. 2011. Diagnosing tuberculosis with urine lipoarabinomannan: systematic review and meta-analysis. Eur Respir J 38:1398–1405.

166. Sinkov V V., Ogarkov OB, Plotnikov AO, Gogoleva NE, Zhdanova SN, Pervanchuk VL, Belkova NL, Koshcheev ME, Thomas TA, Liu J, Zorkaltseva EY, Heysell SK. 2019. Metagenomic analysis of mycobacterial transrenal DNA in patients with HIV and tuberculosis coinfection. Infect Genet Evol 104057.

167. Gansauge M-T, Meyer M. 2013. Single-stranded DNA library preparation for the sequencing of ancient or damaged DNA. Nat Protoc 8:737–48.

168. Troll CJ, Kapp J, Rao V, Harkins KM, Cole C, Naughton C, Morgan JM, Shapiro B, Green RE. 2019. A ligation-based single-stranded library preparation method to analyze cell-free DNA and synthetic oligos. BMC Genomics 20:1023.

169. Aronesty E. 2013. Comparison of Sequencing Utility Programs. Open Bioinforma J 7:1–8.

170. Wood DE, Lu J, Langmead B. 2019. Improved metagenomic analysis with Kraken 2. Genome Biol 20:1–13.

171. Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM.

172. Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. deepTools2: a next generation web server for deep-sequencing data analysis. Nucleic Acids Res 44:W160–W165.

173. Pearce MM, Hilt EE, Rosenfeld AB, Zilliox MJ, Thomas-White K, Fok C, Kliethermes S, Schreckenberger PC, Brubaker L, Gai X, Wolfe AJ. 2014. The female urinary microbiome: A comparison of women with and without urgency urinary incontinence. MBio 5:e01283-14.

174. Stoddard SF, Smith BJ, Hein R, Roller BRK, Schmidt TM. 2015. rrnDB: Improved tools for interpreting rRNA gene abundance in bacteria and archaea and a new foundation for future development. Nucleic Acids Res 43:D593–D598.

175. Clarridge JE. 2004. Impact of 16S rRNA gene sequence analysis for identification of bacteria on clinical microbiology and infectious diseases. Clin Microbiol Rev. American Society for Microbiology (ASM).

176. Collins D. 1991. Identification of an insertion sequence, IS1081, in Mycobacterium bovis. FEMS Microbiol Lett 83:11–15.

177. van Soolingen D, Hermans PW, de Haas PE, van Embden JD. 1992. Insertion element IS1081-associated restriction fragment length polymorphisms in Mycobacterium tuberculosis complex species: a reliable tool for recognizing Mycobacterium bovis BCG. J Clin Microbiol 30.

178. Beggs ML, Cave MD, Marlowe C, Cloney L, Duck P, Eisenach KD. 1996. Characterization of Mycobacterium tuberculosis complex direct repeat sequence for use in cycling probe reaction. J Clin Microbiol 34:2985–2989.

179. Li W, Freudenberg J. 2014. Mappability and read length. Front Genet 5:1–1.

180. Chakravorty S, Simmons AM, Rowneki M, Parmar H, Cao Y, Ryan J, Banada PP, Deshpande S, Shenai S, Gall A, Glass J, Krieswirth B, Schumacher SG, Nabeta P, Tukvadze N, Rodrigues C, Skrahina A, Tagliani E, Cirillo DM, Davidow A, Denkinger CM, Persing D, Kwiatkowski R, Jones M, Alland D. 2017. The new Xpert MTB/RIF ultra: Improving detection of Mycobacterium tuberculosis and resistance to Rifampin in an assay suitable for point-of-care testing. MBio 8.

181. Boyle DS, McNerney R, Teng Low H, Leader BT, Pérez-Osorio AC, Meyer JC, O'Sullivan DM, Brooks DG, Piepenburg O, Forrest MS. 2014. Rapid detection of Mycobacterium tuberculosis by recombinase polymerase amplification. PLoS One 9.

182. Nghiem MN, Van Nguyen B, Nguyen ST, Vo TTB, Van Nong H. 2015. A simple, single triplex PCR of IS6110, IS1081, and 23S ribosomal DNA targets, developed for rapid detection and discrimination of mycobacterium from clinical samples. J Microbiol Biotechnol 25:745–752.

183. Khosravi AD, Alami A, Meghdadi H, Hosseini AA. 2017. Identification of Mycobacterium tuberculosis in Clinical Specimens of Patients Suspected of Having Extrapulmonary Tuberculosis by Application of Nested PCR on Five Different Genes. Front Cell Infect Microbiol 7:3.

184. Pau C-P, Wells SK, Rudolph DL, Owen SM, Granade TC. 2010. A rapid real-time PCR assay for the detection of HIV-1 proviral DNA using double-stranded primer. J Virol Methods 164:55–62.

185. Latorra D, Campbell K, Wolter A, Hurley JM. 2003. Enhanced allele-specific PCR discrimination in SNP genotyping using 3? locked nucleic acid (LNA) primers. Hum Mutat 22:79–85.

186. Chen C, Ridzon DA, Broomer AJ, Zhou Z, Lee DH, Nguyen JT, Barbisin M, Xu NL, Mahuvakar VR, Andersen MR, Lao KQ, Livak KJ, Guegler KJ. 2005. Real-time quantification of microRNAs by stem-

loop RT-PCR. Nucleic Acids Res 33:1–9.

187. Guescini M, Sisti D, Rocchi MBL, Stocchi L, Stocchi V. 2008. A new real-time PCR method to overcome significant quantitative inaccuracy due to slight amplification inhibition. BMC Bioinformatics 9:326.

188. Neto MF, Butzler MA, Reed JL, Rui X, Fisher MJ, Kelso DM, McFall SM. 2017. Immiscible phase filter extraction and equivalent amplification of genotypes 1–6 of hepatitis C RNA: The building blocks for point-of-care diagnosis. J Virol Methods 248:107–115.

189. Creecy A, Russ PK, Solinas F, Wright DW, Haselton FR. 2015. Tuberculosis Biomarker Extraction and Isothermal Amplification in an Integrated Diagnostic Device. PLoS One 10:e0130260.

190. Guckenberger DJ, Pezzi HM, Regier MC, Berry SM, Fawcett K, Barrett K, Beebe DJ. 2016. Magnetic System for Automated Manipulation of Paramagnetic Particles. Anal Chem 88:9902–9907.

191. Berry SM, Pezzi HM, LaVanway AJ, Guckenberger DJ, Anderson MA, Beebe DJ. 2016. AirJump: Using Interfaces to Instantly Perform Simultaneous Extractions. ACS Appl Mater Interfaces 8:15040–15045.

192. Pearlman SI, Leelawong M, Richardson KA, Adams NM, Russ PK, Pask ME, Wolfe AE, Wessely C, Haselton FR. 2020. Low-Resource Nucleic Acid Extraction Method Enabled by High-Gradient Magnetic Separation. ACS Appl Mater Interfaces 12:12457–12467.

193. Creager RS, Irvine B, Le I WA. 2019. nRichDX Revolution Instrument and cfDNA Isolation Kit for Extraction of cfDNA from Large Plasma and Urine Sample Volumes Improves Yield of Rare Targets, abstr TT039. J Clin Microbiol 21:1239.