Development of an amplicon-based NGS protocol for HIV-1 subtype B drug-resistance surveillance using degraded samples

by

Kurt Patrick Kolsun

A Thesis submitted to the Faculty of Graduate Studies of

The University of Manitoba

in partial fulfilment of the requirements of the degree of

MASTER OF SCIENCE

Department of Microbiology, Faculty of Science

University of Manitoba

Winnipeg, Manitoba, Canada

Abstract

With the widespread use of antiretroviral therapy, HIV drug resistance (HIVDR) rates are increasing globally. Many reported cases of HIV are found in areas without readily accessible testing centres capable of processing blood samples. Blood samples received from these areas can become degraded due to storage conditions during transportation to another facility or country capable of testing. Degraded samples present challenges for HIVDR testing, due to compromised viral RNA integrity interfering with RNA amplification and sequencing. Similarly, archival samples and dried blood spots (DBS) can also pose challenges. DBS samples are a highly accessible and cost-effective sample collection method used for HIVDR testing in resource-limited settings within Canada and across the world; however, they can be difficult to process through traditional methods due to susceptibility to degradation and low input volumes. To address this, we have developed a multiplex, amplicon-based, direct-to-sequencing HIVDR testing protocol compatible with degraded samples. This approach uses a hemi-nested approach with primer pools spanning the entire *pol* region of the HIV-1 genome. Protocol primers also incorporated Illumina P5/P7 adaptors, single index barcodes, and unique molecular identifiers (UMIs or Primer ID). This approach allow direct sequencing of amplicons, decreasing overall cost, mitigating PCR biases, and increasing clinical throughput. Here, we show the utility of this approach, successfully targeting the entire *pol* gene region of the HIV-1 subtype B genome in degraded and DBS samples. Preliminary work suggests the protocol may be adaptable for other major subtypes circulating in Canada as data shows the primer pools can amplify other major HIV-1 subtypes, albeit this may require further optimization. This protocol provides a potential option for NGS-based HIV drug-resistance testing from degraded samples and phylogenetic studies, while also building towards the ultimate goal of increasing accessibility of HIV diagnostics for all Canadians.

Acknowledgements

To start, I would like to give a big thank you to my supervisor, Dr. Michael Becker. Michael, thank you for providing me the opportunity to pursue my master's degree at a federal government laboratory and for taking a chance on me as a new supervisor within the Department of Microbiology. I truly appreciate all of the support, guidance, and opportunities that you have provided me throughout my studies; I have become a better scientist because of you. I would also like to thank my co-advisor, Dr. Silvia Cardona, along with my committee members, Dr. Hezhao Ji and Dr. Aleeza Gerstein, for their input, consultation, and mentorship while completing my research.

To everyone at the JC Wilt Infectious Diseases Research Centre who I got to connect with, learn from, and build friendships with, thank you. Your kind and welcoming personalities truly helped create an environment where I felt comfortable being myself. A special thank you to those from the Dr. Michael Becker Lab including Leandro Crisostomo, Branden Gregorchuk, Shayna Giesbrecht, Hannah Exner, Sara Kost, Roniele Cordeiro, and Breanne Head. The amount of support, advice, and friendship you all provided to me is invaluable and I will cherish all of the laughs we have shared together. To the two undergraduate students that I mentored in the lab, Katia Messner and Asad Nasir, thank you for always being inquisitive, asking me thoughtful questions, and working extremely hard. As well, a special thank you to Rupert Capina from the National Laboratory for HIV Genetics, and to "The Anchormen" Chantal Munyuza, Alexandria Reimer, and Rayeil Chua. Thank you all for always greeting me with a smile, teaching me new concepts by answering every question I had, and providing meaningful connections that I am truly grateful for.

Lastly, I would like to thank my parents, Wendell and Sheila, who have provided me with an immeasurable amount of support throughout my entire scholastic career, especially during my graduate studies. They have always been my biggest supporters, teaching and encouraging me to never give up even when things get really hard. I also want to thank my girlfriend, Kamila, for always being by my side and supporting me through the many highs and lows of graduate studies. Thank you for always letting me talk to you about my research, or practice a presentation in front of you, even though you didn't know what I was talking about most of the time.

Table of Contents

Author Contributions

FSWEP undergraduate students, Katia Messner and Asad Nasir, contributed to experimentation, including PCR and sample purification. Chantal Munyuza assisted in Nextera XT DNA Library preparation kit with the Illumina MiSeq using National Microbiology Laboratory standard operating procedures. Chantal Munyuza and Alexandria Reimer assisted in extraction of certain dried blood spot samples using National Microbiology Laboratory standard operating procedures. While these individuals helped me perform a subset of experiments across iterations of the protocol, they worked under my supervision. Ultimately, I was responsible for experimental design, data analysis, and interpretation of the results.

List of Tables

List of Figures

List of Abbreviations

AIDS	Autoimmune deficiency syndrome
ART	Antiretroviral therapy
CBPH	Capture-based probe hybridization
CRF	Circulating recombinant form
DBS	Dried blood spots
DNA	Deoxyribonucleic acid
DRM	Drug resistance mutation
EQAPOL	External quality assurance program oversight laboratory
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
HIVDR	Human immunodeficiency virus drug resistance
IN	Integrase
LADRV	Low-abundance drug resistance variants
LTR	Long-terminal repeat
NGS	Next-generation sequencing
NML	National Microbiology Laboratory
NRTI	Nucleoside reverse transcriptase inhibitor
NNRTI	Non-nucleoside reverse transcriptase inhibitor
PCR	Polymerase chain reaction
PR	Protease
RT	Reverse transcriptase
RNA	Ribonucleic acid
UMI	Unique molecular identifier

Chapter 1: Literature Review

1.1 Human Immunodeficiency Virus (HIV)

Human immunodeficiency virus (HIV) is one of the most extensively studied RNA viruses that has no known cure or vaccine, and is of one of the largest, ongoing epidemics in human history [1–3]. Since the beginning of the HIV epidemic, approximately 85.6 million people have been infected with HIV, with approximately 40.4 million (32.9-51.3 million) having died of AIDS-related complications [1,3]. At the end of 2022, it was estimated that 39 million (33.1-45.7 million) people globally were living with HIV [1]. In Canada, an estimated 62,790 people were living with HIV at the end of 2020 [4], with an additional 1,722 newly diagnosed cases in 2021 [5] and 1,833 newly diagnosed cases in 2022 [6].

HIV is a part of the genus Lentivirus within the Retroviridae family and subfamily Orthoretrovirinae [7,8]. HIV is classified into two different types based on their genetic characteristics and viral antigenic differences [7]: type 1 (HIV-1) and type 2 (HIV-2). As a result of current epidemiologic and phylogenetic analyses, HIV is extrapolated to have been introduced into the human population around 1920 to 1940 [7,9,10]. Due to the close evolutionary relationship between simian immunodeficiency virus (SIV) and HIV, HIV-1 was determined to have evolved from non-human primate immunodeficiency viruses from West Central African chimpanzees (SIVcpz) and HIV-2 from West African sooty mangabeys [10].

HIV is transmitted primarily through bodily fluids and targets cells that contain CD4 receptors, which are essential to the viral particle's recognition and entry into the host cell [7]. CD4-containing cells include macrophages, T-helper cells, dendritic cells and astrocytes [7]. Since HIV infections are lifelong, continual degradation of CD4-containing cells weakens the immune

system over time, making the individual more susceptible to additional infections and diseases, potentially leading to aids-related complications and death [7].

However, it is possible to delay the progression of an HIV infection into AIDS through the use of highly active antiretroviral therapy (HAART), which involves three to four drugs, each from a different class of HIV antiretroviral drugs [2,7]. Treating HIV-positive individuals through the use of HAART can help to improve their overall quality of life. The development of HAART treatment was a product of many years of research and is an example of the importance of HIV research in improving the lives of patients. Furthermore, HIV research has led to the implementation of increased safety requirements in the production of blood components and plasma derivatives to ensure the transmission of viruses does not occur through these avenues [7].

Even with the success of HAART and improvements in blood and plasma component safety, the primary goal in eliminating HIV is still unresolved. Currently, there is no HIV vaccine that has been approved for use [7]. There is also the ongoing struggle with HIV drug-resistance (HIVDR) that has surfaced due to the widespread use of HAART treatments [2]. Therefore, continued work is imperative to developing a vaccine, improving the livelihood of HIV-positive individuals with better treatment options, and monitoring the spread, evolution, and emergence of drug resistant HIV strains.

1.2 Human Immunodeficiency Virus Type 1 (HIV-1)

HIV-1 is subdivided into groups M (main/major), N (non-M, non-O), O (outlier), and P with the main (M) group, infecting at least 60 million people and causing around 25 million deaths [8,11]. HIV-1 group M viruses are further subdivided into subtypes A to D, F to H, J and K, with groups A and D evolutionarily being the oldest viruses and subtypes B and D being very closely

related and considered sub-subtypes [7,8]. HIV-1 accounts for 95% of infections globally [12], with HIV-1 group M subtype B being of particular interest to those in North America, as it is the most prominent circulating HIV-1 subtype [13]. There are also recombinant strains of HIV-1 that carry regions from two genetically distinct parental HIV-1 strains, which are called circulating recombinant forms (CRFs) [7,13]. CRF formation depends on the combination of viral genomes of different subtypes affecting the same cell, which leads to multiple changes in the genome and results in genome regions from each subtype found in the CRF [8,13].

The HIV-1 genome consists of two identical, single-stranded RNA molecules enclosed within the core of the virus particle [14]. The HIV proviral genome, also known as proviral DNA, is generated by the reverse transcription of the viral RNA genome, followed by the integration of double-stranded HIV DNA into the human genome [7,15]. The DNA genome is flanked at either end by long terminal repeat (LTR) sequences, with the 5' LTR region encoding for the promoter for transcription of the viral genes. Following in the 5' to 3' direction, the *gag* gene reading frame follows the 5' LTR and encodes the proteins for the following: outer core membrane (p17), the capsid protein (p24), the nucleocapsid (p7), and a smaller, nucleic acid stabilizing-protein [7,14]. Directly proceeding the *gag* reading frame is the *pol* gene reading frame. The *pol* gene encodes protease (PR, p12), reverse transcriptase (RT, p51), RNase H (p15), and integrase (IN, p32) enzymes [7,14]. After the *pol* gene follows the *env* gene reading frame encoding two envelope glycoproteins, surface gp120 and transmembrane gp41 [7,15]. In addition to structural proteins, the HIV genome encodes other regulatory proteins important for the initiation of HIV replication, such as transactivator protein (Tat) and RNA splicing-regulator (Rev; [7]). Other proteins, such as the negative regulating factor (Nef), viral infectivity factor (Vif), virus protein r (Vpr), and virus protein unique (Vpu), are important for viral replication, virus budding, and pathogenesis [7].

1.3 HIV-1 Drug Resistance

With the high mutation rate of the HIV RT enzyme and widespread use of HAART, arguably the development and identification of drug resistance mutations (DRMs) in circulating HIV strains was inevitable [16]. The development of DRMs, specifically within the *pol* gene, have resulted from a combination of factors, including a high viral replication rate leading to many progeny [17], error-prone replication processes by the RT enzyme [17], and the selective pressure exerted by antiretroviral therapy (ART) drugs [18]. Clinical applications of ART drugs target mainly the *pol* gene, and its encoding regions due to their central role in viral replication and infection. Due to all of these factors, the majority of DRMs known today are found within the *pol* gene, specifically in the PR, RT, and IN encoding regions [19–21]. Although the majority of DRMs are found within the *pol* gene region, other regions of the genome also contain potential DRMs, such as the gp120 and gp41 regions of *gag*, or the *env* gene, which encode proteins that mediate the viral particle's fusion with a CD4-containing host cell [7,17,21]. While the *gag* and *env* gene regions are outside of routine HIVDR surveillance, studies have shown that these regions, specifically within *gag*, can exacerbate the effects of DRMs found within *pol* [22,23] necessitating more broad surveillance strategies.

Within *pol*, four different types of drug classes target the different gene products. RT activity is targeted by two drug classes: nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTIs; [7,17,20]), while PR and IN are targeted by specific protease inhibitors and integrase inhibitors [7,17,19,20]. The genomic regions encoding for the enzyme active sites of RT and PR are most prevalent to drug-resistance mutations [17]. Although RT and PR DRMs are the most common, drug resistance has been found in all HIV drug classes. For integrase inhibitors, some mutations allow the HIV mutant to proceed with integration

into the host genome [17]. For *env* gene, gp120 and gp 41 are targeted by fusion/entry inhibitors, which consists of two clinically approved drugs in this class [21]. Even without the use of HAART, drug resistance has emerged or has naturally been found within strains, therefore it is important to monitor for the emergence of putative drug resistant HIV strains.

There are various resources online for researchers and medical practitioners to identify DRMs, such as the International Antiviral Society-USA (IAS-USA) DRM list and the Stanford University HIV Drug Resistance Database [19–21]. These resources define the consensus amino acid and amino acid position where all major DRMs are located, along with the corresponding mutation and respective drug that is affected. The Stanford HIVDR Database also includes an algorithm where an HIV sample consensus sequence can be uploaded to the HIVdb program [24] to identify levels of resistance to the most commonly used protease, nucleoside, non-nucleoside, and integrase inhibitors [19,20]. Another tool of particular significance for HIVDR analysis is the HyDRA Web application [25]. HyDRA Web processes next-generation sequencing (NGS)-derived HIVDR data and generates a report on all HIVDR mutations found in the *pol* gene regions (PR, RT, IN) by referring to the Stanford HIVDR database algorithms [19,20], along with a consensus sequence at a user-defined threshold, complete nucleotide and amino acid variant frequency reports, and read coverage depth across the *pol* gene [25,26].

1.4 Blood Sample Types used in Genotypic HIV Drug Resistance Testing

1.4.1 Traditional Sample Types and Dried Blood Spots

Plasma is the supernatant remaining after the cellular components found in anti-coagulated blood are removed by centrifugation [27]. With regards to HIVDR testing, plasma is the gold standard sample type most commonly used [28,29], serving all clinical, surveillance, and research HIVDR testing needs [27]. The benefits of using plasma over other blood samples types include:

highest viral RNA concentration out of any sample type, accurate representation of the actively circulating HIV population, maximal recovery of cell-free viral RNA in blood, low RNA degradation coupled with high template integrity, and suitability with various assays [27]. The drawbacks to using plasma are the need for phlebotomy, centrifugation, and low-temperature transportation and storage, as well as poor HIV amplification when viral load is low [27,28].

Serum is the fluid remaining after whole blood is naturally clotted [27]. Serum is nearly identical to plasma in composition, with the exception of fibrinogen, which is present in plasma but naturally removed from serum during clotting [27]. Due to their similarity, serum is the most commonly used substitute for plasma in molecular HIV assays and while plasma is the preferred sample type, serum is often used for serological HIV diagnosis [27]. While similar to plasma, serum has the potential for viral RNA to be trapped in blood coagulum during clotting, contributing to lower viral loads in serum samples [27,30]; however, there is no evidence that this affects HIVDR profiles [27].

While serum and plasma have many benefits, they also have drawbacks, especially in resource limited settings [27,31]. Both require phlebotomy and skilled lab personnel, strict cold-chain transportation, and low-temperature storage conditions to maintain the specimen quality [27]. Additionally, through exclusion of the cellular component when extracting serum or plasma, it may be challenging to detect proviral DNA. Therefore, there is a need for alternative sample types, such as dried blood spots (DBS), when collecting samples for HIVDR surveillance in resource-limited settings [32].

DBS are prepared from whole blood, which has a plasma component, containing circulating HIV-1 particles, and a cellular component, which includes human lymphocytes

containing proviral DNA [27,33,34]. Preparation of DBS entails spotting and drying ~50-75µL of whole blood per spot onto an absorbent filter paper card [27,32,33,35]. DBS is a highly accessible sample type; whole blood can be collected via phlebotomy or from a simple, non-invasive pinprick collected by either a trained professional, trained member of an organization or community, or the patient themselves [27]. DBS is also a practical and cost-effective sample type for HIVDR testing as it can be naturally dried, shipped in a standard envelope through regular post, and stored at ambient temperatures using zip-lock bags with desiccant for days or weeks before being processed [27,32,33,36].

DBS is currently one of the two samples types collected for HIVDR genotyping [28] and well-established DBS-based HIVDR testing guidelines have been implemented [34]. While DBS is considered a suitable alternate sample type for HIVDR testing, it does have intrinsic limitations. Like whole blood, DBS contains proviral DNA [27]. Although proviral DNA enhances HIV amplification and facilitates initial diagnosis, it can also skew the accurate measurement of replication-active HIV viruses circulating within an individual [27]. As a result, proviral DNA may not accurately capture the true HIV drug resistance status of a patient [27]. Due to its convenience and accessibility, DBS may also be susceptible to contamination by the user, environment, or via contact with other DBS cards [35]. Viral RNA in DBS can also become degraded depending on the storage and shipping conditions used to send the samples to a centralized laboratory for HIVDR testing, particularly if the samples are shipped internationally or exposed to moisture [27]. As a result, DBS has been mainly applied in HIVDR surveillance and research studies [27].

1.5 Sequencing Protocols Used for HIVDR Surveillance and Analysis

Two main sequencing techniques for HIV are currently used: Sanger sequencing and NGS-based HIVDR assays. Sanger sequencing is widely used as the clinical diagnostic method for

HIVDR genotyping. The technique involves di-deoxy chain termination chemistry, generating a single consensus sequence of the most common bases at each nucleotide position [37]. When used for HIVDR testing, Sanger sequencing has been able to accurately detect DRMs found at a detection frequency of greater than 20% [38]. However, due to the nature of Sanger sequencing's population-based methodology, it does not reliably detect mutations that are present at a lower frequency (< 20%), also known as low abundance drug resistance variants (LADRVs) within a viral sample [38,39]. As a result, the main driver for introducing NGS-based techniques into HIVDR testing were due to the emergence of LADRVs, and Sanger sequencing's inability to accurately detect mutations lower than 20% frequency. It should also be noted that while batched HIVDR testing, which would be commonly done with NGS-based assays, serves surveillance purposes, individualized HIVDR tests primarily guide clinical patient care. Therefore, Sanger sequencing still very much has clinical relevance for HIVDR testing.

NGS-based HIVDR assays have been recently adopted for HIVDR testing due to their ability to detect LADRVs. As well, while NGS-based assays may vary in chemistries utilized, they all have common characteristics of generating large amounts of output reads and the ability to sequence multiple different samples at the exact same time on a single sequencing run [40]. NGS-based HIVDR assays can also reduce the cost of clinical HIVDR testing due to having a lower cost per test through sample pooling compared to Sanger sequencing [38]. While it is a promising new alternative, there are still some caveats with the approach compared to Sanger sequencing, such as NGS-based assays having higher sequencing errors comparatively, more labor-intensive workflows, and longer turnaround times [38]. HIV diagnostics and epidemiology often require working with degraded samples, and therefore, specialized techniques have been developed to address this challenge [7,41,42]. Currently, there are two main amplification approaches used with

high throughput HIV sequencing: amplicon-based sequencing (also known as amplicon-tiling or "RNA jackhammering") and capture-based probe-hybridization (CBPH). Both of these approaches are targeted, requiring previous knowledge of genome sequence, in contrast to agnostic methods that require little *a priori* genomic information [43]. As both amplicon-based sequencing and CBPH can support full genome sequencing from low titre degraded samples [41,42,44], they are valuable tools that may support robust HIV sequencing protocols. However, both approaches can vary greatly in effectiveness, as they are heavily dependent on the design of the primer/probe sequences [43]. Nonetheless, both CBPH and RNA jackhammering may be more compatible with degraded or low abundance samples compared to traditional methods such as shotgun metagenomics [43,45] or conventional reverse transcription PCR [42].

1.5.1 Amplicon-Based Sequencing

Amplicon-tiling involves the amplification of a target genome, or genome region, using a series of multiplex primer pools [43,46]. In this manner, amplicon-tiling follows the same process as individual amplicon sequencing, except it involves multiple primer pairs hybridizing to various target regions across the genome in a single reaction. Due to the increased number of primers utilized within a single reaction, target genome regions are alternated between primer pools to reduce the amount of overlap between target sites [42,46–48]. Reducing overlap promotes evenness of coverage and prevents incorporation of an internal primer sequence into any given amplicon [42,46]. Using one external primer upstream of the target region, and one internal, target region primer, a longer template molecule is formed. This results in a higher instance of the target amplicon region, and increases the binding likelihood of the internal primers to the target amplicon region [49].

Hemi-nested PCR (or semi-nested PCR) and nested PCR are other common techniques that can be utilized within amplicon-tiling protocols to further increase reaction sensitivity and specificity. Sensitivity refers to the probability that an assay will produce a positive test result if the target sequences of interest (in this case a DRM) is present, thus a high sensitivity yields few false negatives [40]. Specificity refers to the probability that an assay will not detect target sequences of interest if they are absent, thus a high specificity will produce few false positives [40]. In the case of developing molecular testing tools, such as for HIVDR, specificity and sensitivity help researchers determine the efficacy and reliability of tests by providing proportions based on the number of runs that do not produce false positives or false negatives. Tests that are able to provide high specificity and sensitivity proportions, leading to highly accurate testing results, will be selected for clinical diagnostic use. Due to a lack of universally conserved genome markers in RNA viruses, a successful amplicon-tiling protocol requires careful primer design to target conserved regions, such as the *pol* gene region in HIV-1 [43]. Degenerate primers are also relied on heavily to account for viral diversity [50]. For example, within HIV-1 group M, degenerate primer design accounts for sequence variation between all subtypes to maximize successful amplification [42].

Amplicon-tiling has been utilized for detecting various RNA viruses, such as Zika virus [46,51,52], SARS-CoV-2 [53,54], respiratory syncytial virus A and B [55], tick-borne encephalitis virus [45], Hantaan orthohantavirus [56,57], and Seoul virus [58], among others. While it has great utility, the technique is highly specific and requires well characterized viruses that have numerous reference viral genome sequences available during primer design [43]. With respect to HIV, all of these viruses are single-stranded RNA viruses, needing a reverse transcription enzyme to produce their necessary proteins for viral replication. While these viruses may have that commonality, HIV

is particularly challenging to deal with due to its notably error-prone reverse transcription enzyme, extremely complex RNA secondary structure [59], and numerous genotypes, subtypes, and quasispecies. All of these factors make it especially challenging to develop a sequencing protocol for HIV as these differences can affect the PCR amplification of HIV viral RNA, such as the performance of primer binding and efficiency due to various nucleotide differences within genotypes and subtypes. Limited reference sequences pose a barrier, particularly for emerging viruses such as SARS-CoV-2, when there was limited sequencing data available at the start of the epidemic. To overcome this barrier, Quick et al. [46] developed a primer designer software called Primal Scheme, which allows for the efficient development of multiplex primer pair combinations or "schemes" to generate overlapping products across a target genome directly from clinical samples [46]. For example, using Primal Scheme the authors were able to generate an amplicon-tiling method for the entire Zika virus genome called PrimalSeq, which generates 35 overlapping amplicons of ~400 base pairs from two highly multiplexed PCR reactions [60]. This approach is similar to the "jackhammering" method developed for HIV-1 full-genome sequencing [42].

With the large amount of sequencing data now available, many new amplicon-tiling methods have also been developed for SARS-CoV-2. These SARS-CoV-2 amplicon-tiling protocols utilize different approaches, such as using combined short and long amplicon pools [47], singleplex and multiplex amplicon PCR [48], and tailed amplicons [54]. The ARTIC network [61] has also established a method for preparing amplicon pools for SARS-CoV-2 sequencing, which has been used as a standard protocol to assess newly developed amplicon-tiling protocols [47,54]. The latest version of the ARTIC protocol [62] includes coverage of new SARS-CoV-2 variants, utilizes two primer pools conducted in highly multiplexed reactions, and can be sequenced using

either Oxford nanopore technology (e.g. MinION) or Illumina sequencing technology, such as the MiSeq [63].

1.5.2 Capture-Based Probe Hybridization (CBPH)

Amplicon-based sequencing using PCR has advantages such as low cost, rapid processing, automation, sensitivity, and specificity [64]. However due to its high specificity, PCR can be difficult to utilize for highly divergent targets [64]. The competitive nature of multiplex PCR can also introduce biases, favouring shorter amplicons, highly conserved regions, or higher abundance template sequences [65,66]

Capture-based probe hybridization (CBPH) is an approach used to overcome some of these challenges by enriching for specific target sequences [64,67]. This is accomplished through the hybridization of a probes to complimentary genome sequences within prepared libraries [43]. Hybridization can occur on an array, or in-solution through capture of biotinylated probes with streptavidin-coated beads [43,66,67]. The enriched sequencing library may be amplified further, or be used directly for Illumina or Oxford Nanopore sequencing. In the case of Oxford Nanopore sequencing, adaptive sampling, or selective sampling, may further enrich for on-target reads. This process involves the real-time exclusion off-targets during sequencing, by reversing polarity and ejecting them from porin complexes. This process has been described previously for low-abundance species in metagenomics samples [68], and is particularly useful for long read, whole genome sequencing [68]. CBPH can be improved through the selective removal of non-HIV template such as human, bacterial, or other viral sequences [43]. CBPH, when paired with other enrichment methods or selective removal of contaminating sequences, may prove to be a future solution that will promote greater amplification, increased read count, and improved coverage [43]. Currently however, the process is costly, complex, and remains difficult to apply for low viral load

samples. With respect to HIV, CBPH has shown mixed results, producing inconsistent results with low viral load samples below 1,000 copies/mL [44,69,70]. Therefore, we look towards the amplicon-tiling based 'HIV Jackhammer' protocol as the most promising solution for high-throughput clinical testing.

1.6 RNA Jackhammering

Developed by Worobey et al [42], RNA jackhammering was designed to overcome challenges with amplifying and sequencing low template and degraded HIV samples from long-term storage, aiming for amplification of the entire HIV-1 genome. The authors demonstrated that their amplicon-tiling sequencing technique was effective at amplifying small, degraded sequences that were undetected in traditional sequencing approaches. This process, detailed in Figure 1, involves multiple non-overlapping primer pools, that together, cover the entire HIV-1 genome [41,42]. Since HIV and other single-stranded RNA viruses are sensitive to rapid degradation, sequencing of long-term archival HIV samples was nearly impossible prior to the development of this method. With their RNA jackhammering technique, Worobey et al [42] confirmed the route of HIV introduction into North America through sequencing pre-discovery HIV samples of the late 1970s and conducting phylogenetic analyses. With the success of the RNA Jackhammering technique, further investigation into the origin, spread, and overall phylogenetic diversity of HIV is possible. As well, we hypothesize this approach can be useful for tracking the proliferation of anti-viral drug resistance in Canada and globally.

RNA jackhammering involves a multiplex cDNA pre-amplification step, enriching for target sequences ahead of target-specific PCR amplification [41,42]. Ultimately, this improves success rates of the target-specific PCR steps, particularly from low-copy viral samples [42]. Other amplicon-tiling protocols have been inspired by the RNA Jackhammering technique and have been

applied to other viruses such as Zika Virus [46,52,60], and tick-borne encephalitis [45]. For example, Quick et al. [46] tested positive clinical samples and achieved 93% to 99.7% coverage of the entire Zika virus genome using their amplicon scheme sequencing protocol, compared to 0% to 79% genome coverage using metagenomic sequencing approaches.

However, RNA jackhammering does have inherent limitations. One major limitation is that it targets multiple shorter amplicons, which requires a large number of products to generate a tiling path across the HIV genome. Generating amplicons in separate reactions requires an increased number of pipetting steps, further increasing the potential for cross-contamination, greater cost in consumables, and longer processing times [46]. The original RNA Jackhammering protocol conducted reactions on archival frozen serum samples, with the reverse transcription reaction requiring almost three hours to conduct [42]. However, other comparable protocols, such as the PrimalSeq protocol for Zika virus, can conduct the reverse transcription reaction three times faster [46,60]. Another limitation is the amount of optimization needed to optimize the protocol for HIV testing. For example, multiplexing primers can introduce unwanted variables, such as non-specific product formation, along with amplicon generation imbalances due to inherent PCR bias caused by varying primer specificities.

While there are limitations, the RNA jackhammering protocol has been successful amplifying low template, low quality HIV samples [41]. Therefore, this research aims to explore optimization and adaptation of the protocol for routine HIVDR testing of degraded samples. This research can also be leveraged for increasing the phylogenetic understanding of archival HIV-1 subtypes within Canada to determine the origins of transmission. Lastly, the RNA jackhammering technique could be used to identify atypical DRMs found in other areas of the HIV-1 genome outside of the *pol* gene region. Detecting DRMs outside of the *pol* gene would provide more

information to researchers and clinicians interested in DRM surveillance and testing within communities in Canada.

1.7 Current Challenges and Call to Action

In Northern, remote, and isolated communities in Canada, accessible avenues for HIV testing and HIVDR are limited. Due to the limited resources, plasma and serum samples may not be a viable sample collection method, leaving DBS as the best option. Currently, National Microbiology Laboratory (NML) protocols for HIVDR testing require plasma, serum, or DBS sample viral loads of ≥1,000 copies/mL, and mainly target the pol gene region of HIV-1 to monitor for DRMs. Additionally, the current NML sequencing protocol relies on use of the Nextera XT DNA Library Prep kit (Illumina, San Diego, USA), which increases the turnaround time of the protocol to 5 days and the cost per sample to $60. There is a need for the implementation of a sequencing protocol that is cost-effective, compatible with DBS, and capable of identifying DRMs in all areas of the HIV-1 genome.

Taken together, by using a modified approach of the RNA Jackhammering technique, there is the potential to provide complete genomic coverage of the HIV-1 genome, even in heavily degraded or low-copy template samples. Compared to existing approaches, and other potential technologies, this proposed solution is the most likely to improve the current approach to HIVDR clinical testing from DBS and other potentially challenging sample types.

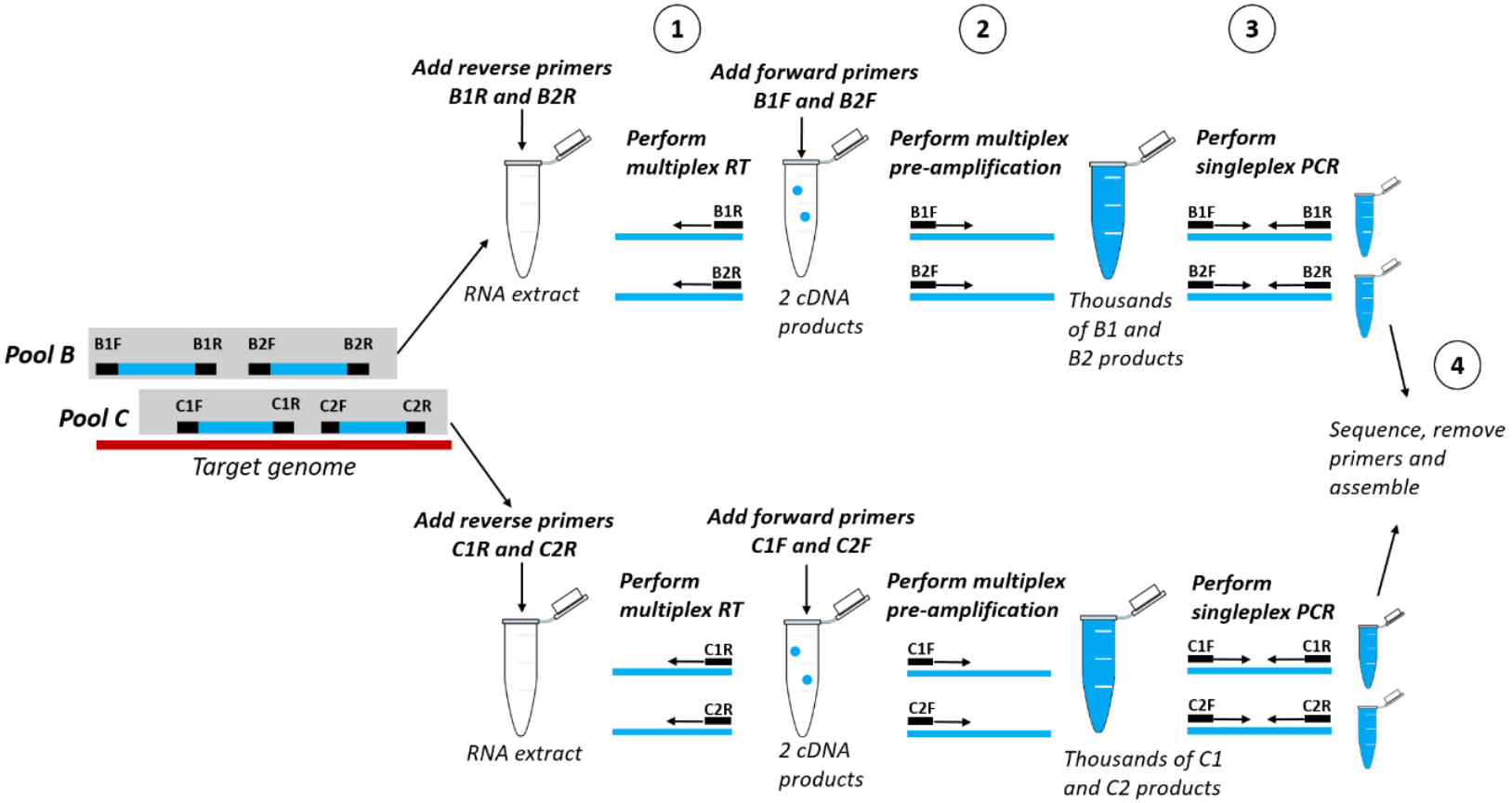

Figure 1. Schematic of the RNA Jackhammering technique. Two non-overlapping primer pools from one primer panel are used to amplify different regions of the HIV-1 genome. In this schematic, only two primer pairs per pool are shown. However, in the largest primer panel of the RNA Jackhammering approach, there are five pools of ten primer pairs. This should ideally lead to the recovery of 10 bands for sequencing and recovery of the HIV-1 genome. Figure adapted from Worobey et al [42].

1.8 Hypothesis

Through extensive optimization, we can adapt the RNA Jackhammering approach for routine sequencing of the HIV-1 subtype B genome from degraded patient samples and dried blood spots. Using this approach, we can likely obtain coverage over the majority of the HIV genome, covering all clinically-relevant mutations underlying HIV drug resistance.

1.9 Objectives

This work aimed to accomplish the following objectives:

1) Validation of previously published methods for whole genome HIV sequencing.

2) Optimization of the protocol for routine clinical surveillance, by investigating aggressive primer pooling strategies (i.e. multiplexing each PCR step within the protocol), adjusting PCR parameters, and automating protocol steps wherever possible.

3) Addition of Illumina adaptors to the protocol primers to allow a direct-to-sequencing approach, as well as the addition of unique molecular identifiers for accurate template quantification.

4) Validation of the optimized protocol using degraded dried blood spots and clinical HIV-1 samples.

Chapter 2: Materials and Methods

2.1 HIV-1 Subtype Samples (gBlocks, Plasma, DBS) and Oligonucleotides

All HIV-1 samples used in this study are described in Table 1. Control oligonucleotides of the HIV-1 genome were designed based on the HIV-1 reference genome (HXB2, subtype B), as well as reference genomes for HIV-1 subtype A, C, and D. All control oligonucleotides were manufactured by Integrated DNA Technologies (IDT, Coralville, USA).

Initial primer sequences used in this study originated from Worobey et al [42]. Primer sequences from the main primer panel (HIVM) in this study are described in Appendix Table A.1. Other primer panels (HIVL and HIVR) used during protocol troubleshooting are described in Appendix Table A.2 and A.3. Alternative HIV-1 primers were designed for areas of low coverage with PrimerDesign-M [71,72] through the Los Alamos National Laboratory HIV Sequence Database (Appendix Table A.4). Potential primer off-targets were identified with Primer-BLAST [73]. All primers were manufactured by Integrated DNA Technologies (IDT, Coralville, USA). Direct sequencing primers were designed based on previously published methods from Zhou et al [74–76] and Thompson et al [77], and were manufactured by Integrated DNA Technologies (IDT, Coralville, USA). Direct sequencing primers were comprehensive, incorporating Illumina P5/P7 adaptors, TruSeq PCR sites, i7 single-index sequences, and unique molecular identifiers (UMIs).

Table 1. List of HIV-1 sample types used in this study

Sample Name		Viral Genome Subtype	Sample Type	Number of Samples	Purpose of sample	Reference or Source
ACCURUN 315 Series 400 HIV-1 RNA positive control		Subtype B	Viral RNA (Plasma)	3	Control	SeraCare (Milford, USA)
Cultured HIV-1 subtype B virus from 8E5 cells (extracted in 2018 & 2021)		Subtype B	Proviral DNA (Plasma)	2	Control	National Microbiology Laboratory Branch
gBlock® synthetic DNA fragments	HIV-1 reference genome (HXB2)	Subtype B Whole genome	Synthetic DNA	1	Control	Integrated DNA Technologies (San Diego, USA) Accession No. K03455.1
	HIV-1 isolate DEMA105TZ001	Subtype A1 Whole genome	Synthetic DNA	1	Control	Integrated DNA Technologies Accession No. JX140650.1
	HIV-1 isolate DEMC09ZA009	Subtype C Whole genome	Synthetic DNA	1	Control	Integrated DNA Technologies Accession No. JX140668.1
	HIV-1 isolate DEMD07UG002	Subtype D Whole genome	Synthetic DNA	1	Control	Integrated DNA Technologies Accession No. KC596071.1
EQAPOL Viral Diversity Panel Samples		Subtype B	Viral RNA (Plasma)	20	Diagnostic	External Quality Assurance Program Oversight Laboratory (Duke Human Vaccine Institute , Durham, USA)

EQAPOL HIV-1 Sample Designation DEMB11US011.01	Subtype B	Viral RNA (Contrived DBS)	6	Degraded	External Quality Assurance Program Oversight Laboratory
Canadian HIV Strain and Drug Resistance (SDR) surveillance program clinical HIV samples	Subtype B	Viral RNA (Serum)	6	Clinical	Public Health Agency of Canada. Update on HIV-1 Strain and Transmitted Drug Resistance in Canada: 2012-2013. Centre for Communicable Diseases and Infection Control, Public Health Agency of Canada, 2017

2.2 Plasma and Serum RNA Extraction

For all plasma samples, including HIV-1 subtype B samples from the External Quality Assurance Program Oversight Laboratory (EQAPOL) Viral Diversity Panel (Duke Human Vaccine Institute, Durham, USA), HIV-1 viral RNA was extracted using the NUCLISENS easyMAG (bioMerieux, Craponne, France). Steps were performed as per the manufacturer's instructions, using Generic Protocol 2.0.1. The protocol entailed on-board lysis with a 10 minute incubation time, a sample input of 450 µL, and elution volume of 100 µL. Extracts were stored at -80°C until used for reverse transcription. For all serum samples, HIV-1 RNA was extracted as described above for plasma with the following modifications: input volume of 400 µL and elution volume of 110 µL.

2.3 Dried Blood Spot Sample Processing, RNA Extraction, and Degradation Study

All DBS samples, including contrived DBS samples created using HIV-1 viral RNA from sample DEMB11US011.01 (EQAPOL), were extracted using the DBS extraction protocol described below. DBS cards were removed from -80°C and incubated at ambient room temperature for 30 minutes before processing. Bitran Series S bags (Globe Scientific, Mahwah, USA) were used for storage of DBS samples. Humidity identification cards within DBS sample storage bags were examined before nucleic acid extraction to ensure that humidity levels did not exceed 30%. When processing DBS samples using a manual DBS puncher, blank DBS cards were punched five times between samples to prevent cross-contamination. For each sample, two blood spots on the DBS card (approximately 150 uL of blood) were punched into a lysis buffer tube. Lysis tubes were rotated for at least two hours. After two hours on the rotisserie, lysis tubes were centrifuged at 1500 RPM for five minutes, and the lysate supernatant was then RNA extracted using the NUCLISENS easyMAG (bioMerieux, Craponne, France). Steps were performed as per the

manufacturer's instructions and in-house standard operating procedure, using protocol Specific B, a sample input of 150 µL, and elution volume of 50 µL. Extracted samples were stored at -80°C until used subsequently for reverse transcription.

For the DBS degradation study, contrived DBS sample cards were placed in an open Bitran Series S bag (Globe Scientific, Mahwah, USA) within a Steri-Cycle CO_2 Incubator (ThermoFisher, Waltham, USA). DBS samples were incubated at 37°C and 91% humidity for various time intervals: 1, 2, 7, 14, and 28-days. As a positive control, a non-degraded DBS sample was extracted and PCR amplified for future comparison to PCR amplification rates of degraded samples using our finalized PCR protocol.

2.4 Amplicon-Tiling of the HIV-1 Genome Using the HIVL and HIVR Primer Panels

Initial validation of the Worobey et al [42] RNA Jackhammering protocol was performed with the HIVL and HIVR primer panels. The HIVL primer panel was designed with larger amplicon targets (~400-600 base pair amplicons) while the HIVR primer panel was designed with smaller amplicon targets (~200-300 base pair amplicons). Figure 2 describes the characteristics each panel and displays the amplicon coordinates of the primer sets across the HIV-1 reference genome (HXB2).

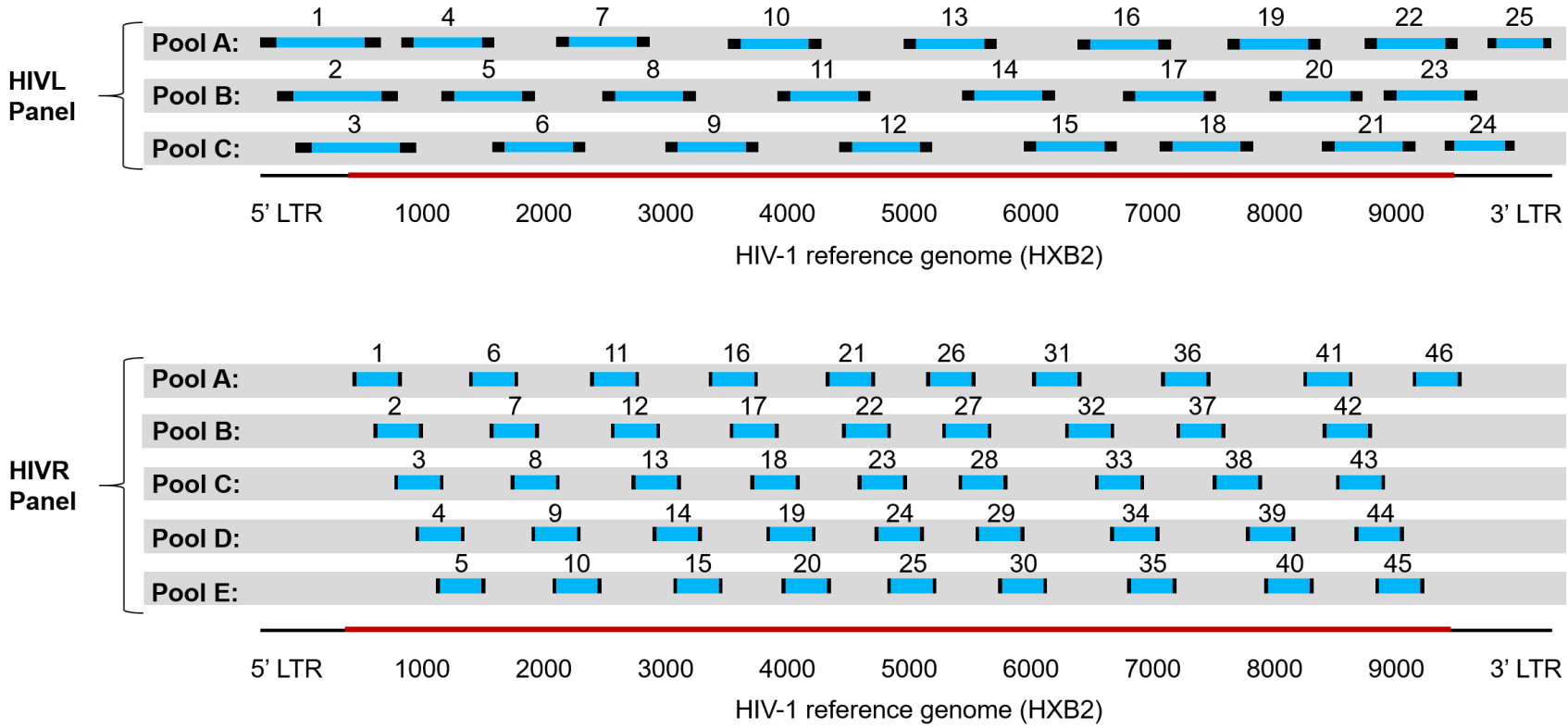

Figure 2. HIVL and HIVR primer panel pool layouts and primer binding locations. HIVL (top) comprises 3 pools of primers (25 primer pairs total) generating longer amplicons of 400-600 base pairs. HIVR (bottom) comprises 5 pools of primers (46 primer pairs total) generating shorter amplicons of 200-300 base pairs. Figure adapted from Worobey et al [42].

2.4.1 HIVL & HIVR Panels - Reverse Transcription

Plasma ACCURUN positive control HIV-1 RNA material (SeraCare, Milford, USA) was extracted using the plasma extraction protocol as described in Section 2.2. ACCURUN positive control RNA (54,000 copies/mL), along with 8E5 HIV-1 proviral DNA (100 copies/µL), were used as starting template material for reverse transcription reactions using the HIVL and HIVR primer panels.

Reverse transcription was performed using the GoScript Reverse Transcription System and GoScript reverse transcriptase (Promega, Madison, USA), as per the manufacturer's recommendations, with an input of 6 µL of RNA template, a pooled primer concentrations of 10 µM of reverse primer, and a final reaction volume of 20 µL. The HIVL primer panel consisted of three individually pooled reactions (8 or 9 primers within each pool) and the HIVR primer panel consisted of five individually pooled reactions (9 or 10 primers within each pool). Incubation conditions for reverse transcription reaction were as follows: denaturation at 70°C for 5 minutes; incubation on ice for 2 minutes; incubation at 50°C for 30 minutes, 55°C for 2 hours, and 85°C for 10 minutes; followed by an indefinite hold at 4°C.

2.4.2 HIVL & HIVR Panels - Pre-Amplification (1st round PCR)

First-strand cDNA was input to a pre-amplification PCR to enrich for target amplicons using the Platinum Taq DNA polymerase High Fidelity PCR kit (Invitrogen, Waltham, USA) as per the manufacturer's recommendations, with an input of 10 µL of amplified template, a pooled primer concentration of 10 µM, and final reaction volume of 50 µL. Pre-amplification reactions for each individual primer pool were conducted in multiplex and added only the forward primers from the respective pool. Incubation conditions were as follows: one cycle of 94°C for 3 minutes;

30 cycles each of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; one cycle of 68°C for 5 minutes; followed by an indefinite hold at 4°C.

2.4.3 HIVL & HIVR Panels - Final Amplification (2nd round PCR)

Pre-amplified PCR products served as template for second round of amplification (final amplification). Final amplification was performed using the Platinum Taq DNA polymerase High Fidelity PCR kit (Invitrogen, Waltham, USA) as per the manufacturer's recommendations, with an input of 10 µL of template, pooled forward and reverse primer concentrations of 10 µM each), and a final reaction volume of 50 µL. Reactions were performed in multiplex within each primer pool, with the following PCR conditions: one cycle of 94°C for 3 minutes; 40 cycles each of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; one cycle of 68°C for 5 minutes; followed by an indefinite hold at 4°C.

2.4.4 HIVL & HIVR Panel Optimization

The original RNA Jackhammering protocol reaction conditions [42] are listed in Table 2 and are compared to our fully multiplexed protocol described throughout Section 2.4. Various steps of our protocol were optimized by adjusting individual reaction variables between different runs. A comprehensive summary of optimized variables is described in Section 3.

Towards optimization of the pre-amplification step, various iterations of the protocol were performed beginning at pre-amplification using 8E5 proviral DNA as the starting template material. Pre-amplification PCR steps were performed for the HIVL and HIVR primer panels at primer pool concentrations of 5µM (0.2 µM final concentration) or 10 µM (0.4 µM final concentration), as well as either 30 or 40 cycles of amplification.

Following validation of the protocol with proviral DNA, protocol optimization shifted back to using HIV-1 viral RNA as the starting template material. Reverse transcription reactions were conducted in singleplex, as well as the preliminary and final amplification PCRs to troubleshoot initial issues with multiplexing the entire PCR protocol. For the HIVL primer panel, the final amplification step was evaluated as a singleplex or multiplex reaction. For the HIVR primer panel, gradient PCR (52°C-57°C) was used to determine the optimal annealing temperature in both the pre-amplification and final amplification PCR steps.

Table 2. Original Worobey et al. "RNA Jackhammering" amplicon-tiling protocol conditions

Condition		RT-PCR Step			
		Reverse Transcription		**Pre-amplification (1st round PCR)**	**Final Amplification (2nd round PCR)**
PCR Kit		GoScript Reverse Transcription System (Promega)		Platinum Taq DNA Polymerase High Fidelity (Invitrogen)	Platinum Taq DNA Polymerase High Fidelity (Invitrogen)
Template Material		HIV-1 viral RNA		1st strand cDNA	Amplified cDNA
# of primers used per PCR		Multiplex		Multiplex	Singleplex
Primer Concentration	Forward Primer (initial; final)	N/A		10 µM (0.4 µM)	10 µM (0.4 µM)
	Reverse Primer (initial; final)	10 µM (1.5 µM)		N/A	10 µM (0.4 µM)
Thermocycler Settings	Annealing Temperature & Time (per cycle)	50°C 30 min	55°C 30 min	52°C	52°C
	Number of Amplification Cycles	4	1	30	40

2.5 Amplicon-tiling of the HIV-1 genome using the HIVM Primer Panel

After initial validation and optimization of the HIVL and HIVR primer panels produced suboptimal results, we shifted focus to the HIVM panel for evaluation. Figure 3 describes the characteristics of the HIVM primer panel and displays the amplicon coordinates across the HIV-1 reference genome (HXB2). While the HIVM and HIVR primer panels are similar in design, there are a couple differences besides the number of amplicons in each panel (46 for HIVR and 50 for HIVM). The HIVM panel contains 10 primer pairs per pool (9 per pool with HIVR, with the exception of Pool A), meaning more primers per multiplex reaction. As a result, the primer binding and target sequence sites vary slightly from those found in the HIVR panel.

2.5.1 HIVM Panel - Reverse Transcription

ACCURUN positive control HIV-1 RNA material (54,000 copies/mL), 8E5 proviral DNA (100 copies/µL), EQAPOL plasma samples (10,000 copies/mL), and contrived DBS samples from EQAPOL DEMB11US011.01 (~45,000 to ~80,000 copies/mL) were all used as starting template material for evaluating reverse transcription reactions involving the HIVM primer panel. Initially, reverse transcription steps utilized the GoScript Reverse Transcription System (Promega, Madison, USA). After protocol optimization, reverse transcription was performed with the Superscript IV first-strand synthesis system (Invitrogen, Waltham, USA). All reverse transcription reactions were conducted as per the manufacturer's recommendations for both kits, with an initial input of either 6 µL or 10 µL of RNA template, a pooled reverse primer concentration of 10 µM, and a final reaction volume of either 20 µL or 50 µL based on the input volume used. The HIVM primer panel consisted of five individually pooled reactions (10 primers within each pool). Reactions for each individual primer pools were conducted in multiplex.

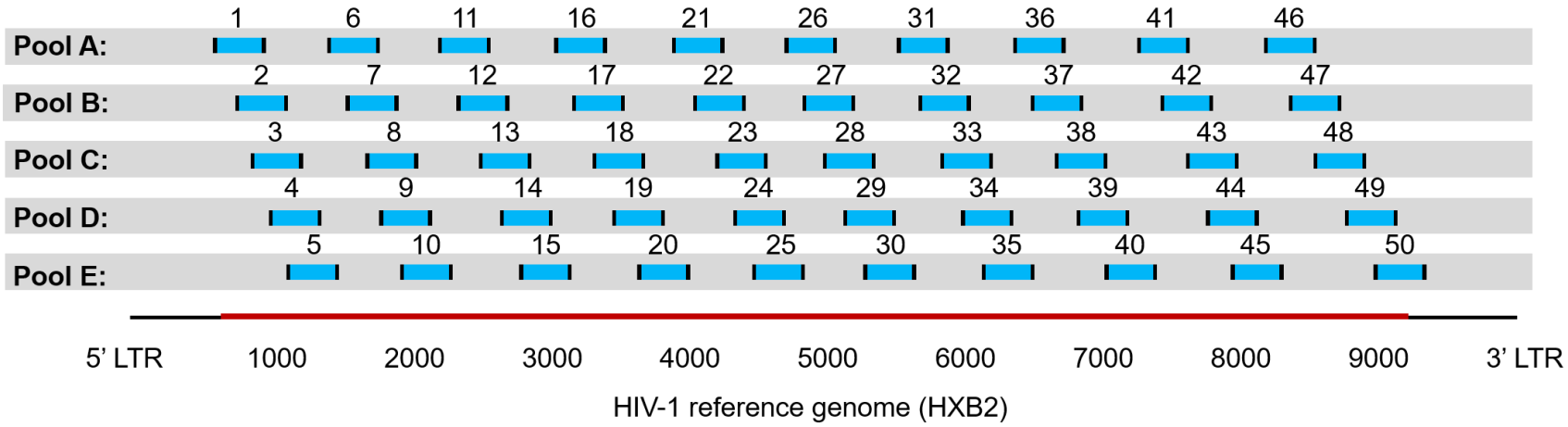

Figure 3. HIVM primer panel pool layouts and primer binding locations. HIVM comprises 5 pools of primers (50 primer pairs total) generating shorter amplicons, anywhere from 200-300 base pairs in size. Figure adapted from Worobey et al. [42].

The GoScript Reverse Transcription System incubation conditions were as previously described with the HIVL and HIVR panels in Section 2.4.1. Incubation conditions for the SuperScript IV first-strand synthesis system were as follows: denaturation at 70°C for 5 minutes; incubation on ice for 2 minutes; incubation at 50°C for 50 minutes, followed by 80°C for 10 minutes; and an indefinite hold at 4°C.

2.5.2 HIVM Panel - Pre-Amplification (1st round PCR)

Pre-amplification PCR was performed in multiplex using the Platinum Taq DNA polymerase High Fidelity PCR kit (Invitrogen, Waltham, USA) as per the manufacturer's recommendations, with a 10 µL input of template from reverse transcription, a pooled forward primer concentration of 10 µM, and a final reaction volume of 50 µL. Forward primers used in each pooled reaction corresponded to reverse primers within the same primer pool used in the RT-PCR as described in Section 2.5.1. Incubation conditions were as follows: one cycle of 94°C for 3 minutes; 30 cycles each of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; one cycle of 68°C for 5 minutes; followed by an indefinite hold at 4°C.

2.5.3 HIVM Panel - Final Amplification (2nd round PCR)

Final amplification was conducted using the Platinum Taq DNA polymerase High Fidelity PCR kit (Invitrogen, Waltham, USA) as per the manufacturer's recommendations, with an input of 10 µL of pre-amplified product, pooled forward and reverse primer concentrations of 10 µM each (final concentration of 0.2 µM per primer pool reaction), and a final reaction volume of 50 µL. Final amplification reactions were conducted in multiplex. Incubation conditions were as follows: one cycle of 94°C for 3 minutes; 40 cycles each of 94°C for 30 seconds, 52°C for 30 seconds, and 68°C for 30 seconds; one cycle of 68°C for 5 minutes; followed by an indefinite hold at 4°C.

2.5.4 HIVM Panel Optimization

Our amplicon tiling PCR protocol using the HIVM primer panel was optimized through the progressive modification of individual reaction variables. Optimization of the pre-amplification step was performed using 8E5 proviral DNA, thereby bypassing reverse transcription. Reverse transcription was optimized using plasma and DBS samples amplification results and adjusting the protocol accordingly for each sample type as previously described. Two different reverse transcription kits were evaluated, the GoScript reverse transcription system and SuperScript IV first-strand synthesis system, across multiple incubation parameters. Multiple reverse primer concentrations were evaluated for reverse transcription including 10 µM, 5 µM, 2.5 µM, 1.75 µM, and 1.0 µM. Pre-amplification was evaluated at final forward primer concentrations of 1 µM, 0.6 µM, or 0.4 µM. Pre-amplification was also evaluated without additional reverse primers, or the addition at a final concentration of 0.4 µM. Lastly, the final amplification PCR was evaluated at final primer concentrations of either 0.4 µM or 0.2 µM. A comprehensive list of all variables tested are further summarized in Section 3 (Results).

To further optimize each reaction for sensitivity and specificity, several PCR techniques were evaluated. For each approach, gradient PCR at 52°C, 55°C, and 57°C was used to determine the optimal annealing temperatures for the pre-amplification and final amplification. The traditional PCR described in Worobey et al. was compared to both hemi-nested and nested PCR strategies. For nested PCR, flanking primers outside the target area of interest in the pre-amplification PCR, and inner primers were subsequently used during the final amplification PCR [49]. The hemi-nested PCR approach utilized an outer forward primer in the pre-amplification PCR but the reverse primer was conserved across both pre-amplification and the final amplification steps [49].

2.6 Evaluation of Optimized Protocol Using Clinical and Contrived Specimens

The final, optimized protocol was tested on EQAPOL plasma samples, contrived DBS samples, and clinically positive HIV-1 subtype B serum samples from the Canadian HIV Strain and Drug Resistance (SDR) Surveillance Program [78]. Nuclease-free water was used as a negative control for both plasma and serum samples. DBS cards containing HIV-negative whole blood were used as negative controls for DBS samples. A complete list of all primers used in the finalized protocol for sequencing the *pol* gene are found in Table A.5.

All reverse transcription reactions were conducted using the SuperScript IV first-strand synthesis system as per the manufacturer's recommendations, with 10 µL of RNA template added, a pooled reverse primer concentration of 10 µM, and a final reaction volume of 50 µL, as per the incubation conditions described in Section 2.5.1. Individual reverse transcription reactions for each primer pool were conducted in multiplex targeting specifically the *pol* region. The resulting first-strand cDNA product was purified and size-selected using Agencourt AMPure XP beads (Beckman Coulter, Brea, USA) as per the manufacturer's protocol. Right-sided cleanup was performed with an AMPure XP bead-to-sample ratio of 1.8x and automated using the epMotion 5075 liquid handling system (Eppendorf, Hamburg, Germany).

Pre-amplification PCR was performed in multiplex using the Platinum Taq DNA polymerase High Fidelity PCR kit (Invitrogen, Waltham, USA) as per the manufacturer's recommendations, with a 10 µL input of template from reverse transcription, a pooled forward primer concentration of 10 µM, a universal reverse primer of 0.4 µM, a final reaction volume of 50 µL, and the same incubations as described for this kit in Section 2.5.2. Pre-amplification was hemi-nested, with forward primers located upstream of the *pol* region target sequences (Figure 3). The resulting first-strand cDNA product was purified and size-selected using Agencourt AMPure

XP beads as per the manufacturer's protocol. Right-sided cleanup was performed with an AMPure XP bead-to-sample ratio of 1.0x and automated using the epMotion 5075 liquid handling system.

The final amplification was conducted in singleplex using the Platinum Taq DNA polymerase High Fidelity PCR kit as per the manufacturer's recommendations, with an input of 10 µL of pre-amplified product, forward and reverse primer concentrations of 0.4 µM, and a final reaction volume of 50 µL. Touchdown-PCR was used, with incubation conditions as follows: denaturation at 94°C for 3 minutes; 10 cycles each of 94°C for 30 seconds, 62°C for 1 minute with step downs of 1.0°C each cycle, and 68°C for 30 seconds; 30 cycles each of 94°C for 30 seconds, 52°C for 1 minute, and 68°C for 30 seconds; one cycle of 68°C for 5 minutes; followed by an indefinite hold at 4°C. As part of the protocol, pol region DRMs that were not fully captured with sufficient coverage were reprocessed, beginning with stored cDNA and specifically targeting any coverage gaps.

2.7 Next-Generation Sequencing Protocol Using Illumina MiSeq

2.7.1 Amplicon Sequencing with the Nextera XT Library Preparation Kit

Sequencing runs using the Nextera XT DNA library preparation kit (Illumina, San Diego, USA) were conducted according to the manufacturer's protocol [79] and automated with the epMotion 5075 liquid handling system (Eppendorf, Hamburg, Germany).

2.7.2 Development, Design, and Implementation of Direct Sequencing Primers

Direct sequencing primers were designed based on the methods of Zhou et al [74–76] and Thompson et al. [77]. UMIs were utilized at the reverse transcription step to facilitate PCR bias corrections during data analysis. Each UMI consisted of a 10-base pair random nucleotide sequence incorporated into reverse transcription primers, as per the methods of Zhou et al [74–76]. Primer sets for pre-amplification and final amplification incorporated Illumina adaptors and

indices to facilitate direct sequencing of final products. Specifically, reverse primers in pre-amplification contained a portion of the Illumina P7 adaptor sequences. Elongated forward primers for the final amplification step included a P5 adaptor sequence, and a reverse primer contained i7 single-index barcodes to facilitate multiplex sequencing. A partial list of direct sequencing primers used for sequencing the *pol* gene region is shown in Table 3 and the full list of primers used can be found in Appendix Tables A.4 – A.7.

2.7.3 Preparation of Illumina Sequencing Libraries Using Custom i7 Primer Constructs

Final amplicon libraries were quantified using the Qubit 1x dsDNA high sensitivity assay kit (Invitrogen, Waltham, MA) and pooled at equimolar concentrations. These equimolar amplicon pools were quantified again via Qubit, before being combined with other samples with unique indices at equimolar concentrations. The final libraries consisting all amplicons and samples to be sequenced was purified and size-selected using Agencourt AMPure XP beads as per the manufacturer's protocol (Beckman Coulter, Brea, USA). Right-sided cleanup was performed with an AMPure XP bead-to-sample ratio of 0.5x, and left-sided cleanup with a ratio of 0.85x. This double-sided cleanup was done to effectively size-select for amplicons that were around 350-550 bp in size, removing any fragments that were larger, or smaller, than this range. Effectiveness of size selection was assessed with the Agilent TapeStation using D1000 ScreenTapes (Agilent, Santa Clara, USA) and the purified mixture was quantified using the Qubit 1x dsDNA high sensitivity assay kit. The molarity of the purified sequencing library was calculated based on the average values from TapeStation and Qubit, and libraries were diluted to 4 nM. Libraries were prepped for sequencing per the Illumina MiSeq system Denature and Dilute Libraries Guide, following Protocol A and a 10% PhiX spike-in [80].

Table 3. List of designed option #1 direct sequencing primer sequences targeting the *pol* gene region.

Primer[a]	5' – 3' Sequence	Comment
HIVM_D2R[b]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GG CYTTTAGTTGCCCYCCTA	cDNA primer with UMI for PR
HIVM_D3R[b]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GG TGYYGCCCTATTTCTAAGTC	cDNA primer with UMI for RT
HIVM_D4R[b]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GG CTRTTRGCTGCCCCATC	cDNA primer with UMI for RT
HIVM_D5.4R[b]	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GG TGCTATTATRTCTAYTATTCTTTCC	cDNA primer with UMI for IN
HIVM_C2F	TGTCAGGGAGTRGGRGGA	1st round PCR forward primer for PR
HIVM_C3F	TGGGCCTGAAAAYCCAT	1st round PCR forward primer for RT
HIVM_C4F	GGAGTGTATTATGACCCATC	1st round PCR forward primer for RT
HIVM_C5F	YCCAGGAATATGGCAAYTAG	1st round PCR forward primer for IN
Universal Adapter Reverse	GTGACTGGAGTTCAGACGTGTGCTC	1st round PCR reverse primer
HIVM_D2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGWTGYACTGAGAGACAGGCTA	2nd round PCR forward primer for PR w/ Illumina adapter sequence

HIVM_D3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGGTAYTGGATGTRGGTGATG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_D4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTTAACAGAGGCAGTGCA	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_D5.4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTAGCAGGAAGATGGCCAG	2nd round PCR forward primer for IN w/ Illumina adapter sequence
Indexed Adapter[c]	CAAGCAGAAGACGGCATACGAGAT**NNNNNN**GTGACTGGAGTTCAG ACGTGTGCTC	2nd round PCR reverse primer w/ Illumina adapter sequence and barcodes

PR = protease, RT = reverse transcriptase, IN = integrase

[a]List of designed direct-sequencing primers is not exhaustive; the sequence table is only a portion of the total. Refer to Appendix Tables A.6 to A.8 for a complete list of primers tested. Primers used in the finalized *pol* gene sequencing protocol are found in Table A.5.

[b]cDNA primer contains a block of degenerate nucleotides (Ns) as the Primer ID.

[c]Indexed adaptors contain Ns which are a set of 48 pre-designed barcodes for multiplexing samples in one run.

2.8 Bioinformatics Processing of Next-Generation Sequencing Data

The bioinformatics workflow used for processing sequencing data generated from Nextera XT libraries is summarized in Figure 4. A schematic of the entire bioinformatics workflow used for processing sequencing data from custom direct-to-sequencing primers is displayed in Figure 5. Raw quality characteristics were assessed with FastQC [81]. For reads including UMIs, UMI-tools *extract* [82] was used to group paired reads based on the 10-base UMI in Read 2 sequences. Following this, reads were trimmed using Trimmomatic [83] to remove adaptor and primer sequences. Trimmed reads were aligned to a reference HIV-1 genome HXB2 (NCBI reference: K03455.1) using bowtie2 [84]. For samples containing UMIs, the BAM output file was processed with UMI-tools *dedup* [82] to isolate unique template reads. Samtools [85] was used to generate DEPTH and FASTQ files from BAM files. DEPTH output files were visualized with the R package *ggplot2* [86] to generate coverage plots. FASTQ files were uploaded into HyDRA [25] for drug-resistance mutation analyses.

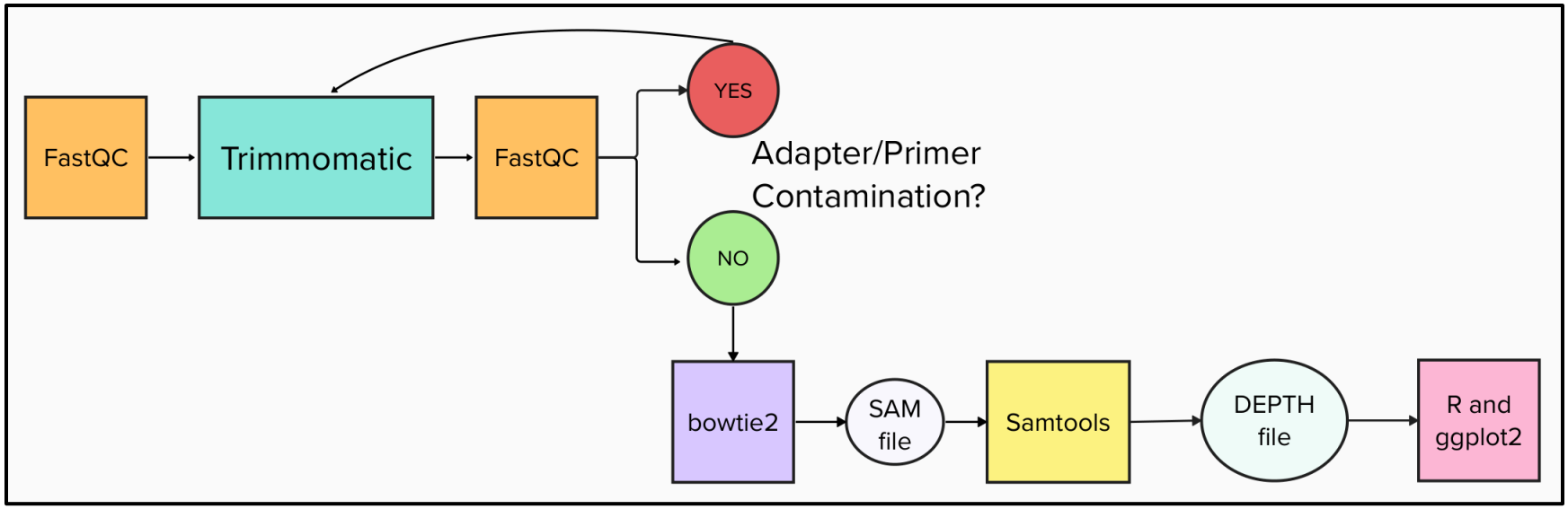

Figure 4. Bioinformatics workflow for processing NGS data without UMIs. FastQC checks for quality metrics of the generated reads and the presence of internal primer or adapter sequences. Trimmomatic removes these unwanted sequences from reads. Bowtie2 performs reference mapping to align reads to the HIV-1 reference genome (HXB2). Samtools converts file formats from a SAM file to a DEPTH file. The resulting DEPTH file is uploaded into the programming language R and used as the primary data source by the R package *ggplot2* to generate a coverage plot for depth of coverage analysis. Squares represent programs used within Linux command line and circles represent output files of these programs.

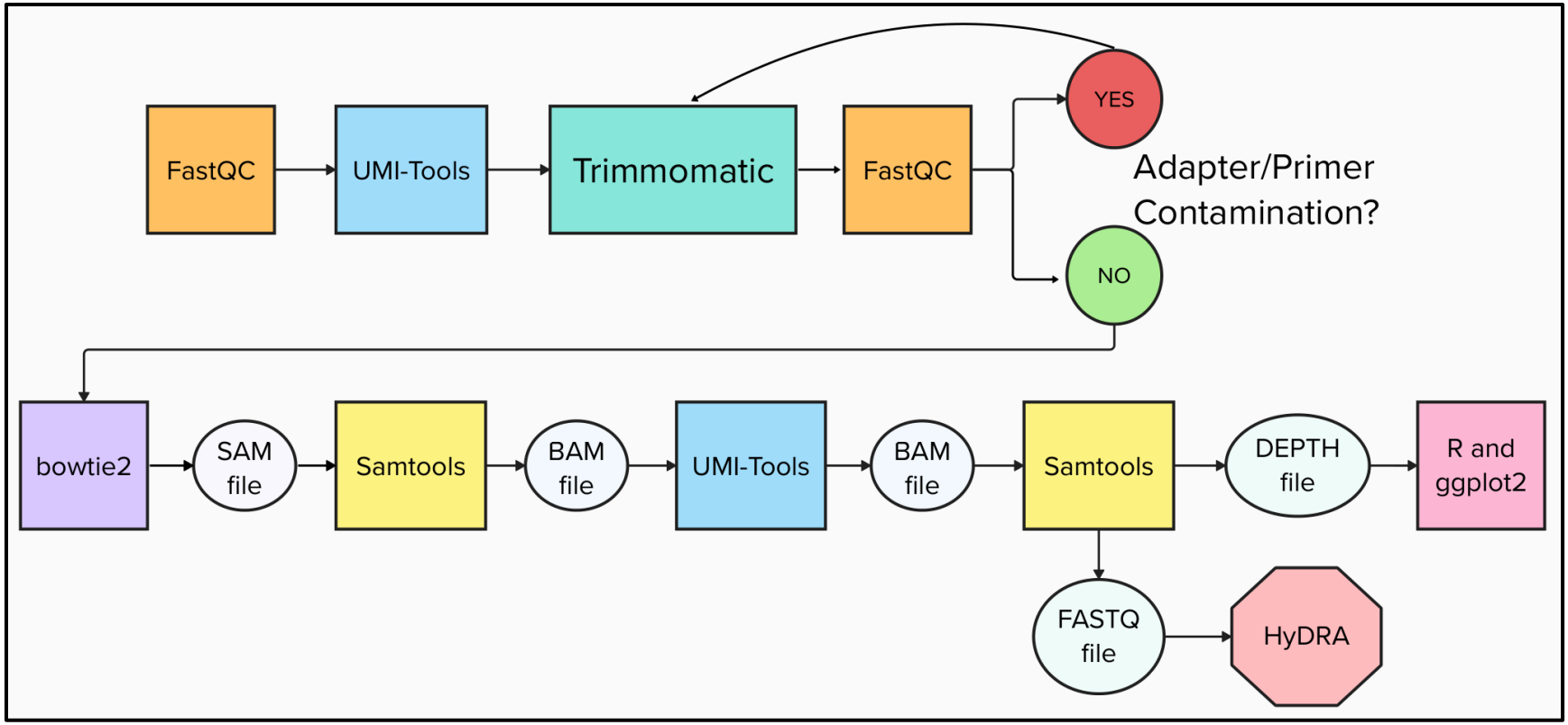

Figure 5. Bioinformatics workflow for processing deduplicated reads using UMIs. FastQC checks for quality metrics of the generated reads and the presence of internal primer or adapter sequences. UMI-tools identifies areas where UMIs are specified in the read sequence and attaches the UMI to the read for further downstream processing. Trimmomatic removes unwanted sequences from reads such as internal primer or adapter sequences. Bowtie2 performs reference mapping to align reads to the HIV-1 reference genome (HXB2). At this point, UMI-tools is the critical tool allowing for recognition and removal of duplicate reads to retain only unique, deduplicated reads. In this pipeline, Samtools converts file formats in one of two ways: 1) from a BAM file to a DEPTH file for upload to R and coverage plot generation and depth of coverage analysis in ggplot2; or 2) from a BAM file to FASTQ file for upload into HyDRA for DRM analysis. Squares represent programs used within Linux command line and circles represent output files of these programs.

Chapter 3: Results

3.1 RNA Jackhammer Protocol Manipulation Using HIVL & HIVR Primer Panels

First, we aimed to replicate the work of Worobey et al. [42], which describes an amplicon-based tiling protocol for HIV-1 full-length genome sequencing (parameters summarized in Table 2 from Chapter 2). Initial experiments utilized primer panels HIVL and HIVR, and initially produced no amplification. Following this, we systematically adjusted PCR parameters including primer concentration, cycling conditions, and PCR kits, as summarized in Table 4 and Table 5. When analyzing amplicon production results, factors such as presence of target amplicon and correct base pair size, band intensity, product yield, and presence (or absence) of off-target bands signaling non-specific binding, directly impacted future optimization decisions. Across all conditions tested, all reactions produced no amplification or non-specific product formation. Due to the limited success of primer panels HIVL and HIVR, we moved forward with the HIVM primer panel (see section 3.2), which provided more robust amplification.

Table 4. List of parameters evaluated during the validation and optimization of the RNA Jackhammering protocol with the HIVL primer panel.

RT-PCR Protocol Step	Condition	Original Condition	Tested Variables	Optimized Condition	Order of Optimization
Reverse Transcription / Pre-amplification	Amplification Starting Point	HIV-1 viral RNA	8E5 proviral DNA	8E5 proviral DNA	3rd
Pre-amplification	Number of Primers per Reaction	Multiplex	Multiplex	Multiplex	1st
			Singleplex		
	Forward Primer Concentration (initial & final)	10 µM (0.4 µM)	5 µM (0.2 µM)	10 µM (0.4 µM)	4th
			10 µM (0.4 µM)		
	Number of Amplification Cycles	30	40	30	5th
Final Amplification PCR	Number of Primers per Reaction	Multiplex	Multiplex	Multiplex	1st
			Singleplex		
	Forward Primer Concentration (initial & final)	10 µM (0.4 µM)	10 µM (0.2 µM)	10 µM (0.4 µM)	2nd
			10 µM (0.4 µM)		
	Reverse Primer Concentration (initial & final)	10 µM (0.4 µM)	10 µM (0.2 µM)	10 µM (0.4 µM)	
			10 µM (0.4 µM)		

Table 5. List of parameters evaluated during the validation and optimization of the RNA Jackhammering protocol with the HIVR primer panel.

RT-PCR Protocol Step	Condition	Original Condition	Tested Variables	Optimized Condition	Order of Optimization
Reverse Transcription / Pre-amplification	Amplification Starting Point	HIV-1 viral RNA	8E5 proviral DNA	8E5 proviral DNA	4th
Pre-amplification	Number of Primers per Reaction	Multiplex	Multiplex	Multiplex	1st
			Singleplex		
	Forward Primer Concentration (initial & final)	10 µM (0.4 µM)	5 µM (0.2 µM)	10 µM (0.4 µM)	5th
			10 µM (0.4 µM)		
	Annealing Temperature and Time (per cycle)	52°C 30 seconds	52°C 30 seconds	52°C 30 seconds	3rd
			55°C 30 seconds		
			57°C 30 seconds		
	Number of Amplification Cycles	30	40	30	6th
Final Amplification PCR	Number of Primers per Reaction	Multiplex	Multiplex	Multiplex	1st
			Singleplex		
	Forward Primer Concentration (initial & final)	10 µM (0.4 µM)	10 µM (0.2 µM)	10 µM (0.4 µM)	2nd
			10 µM (0.4 µM)		
	Reverse Primer	10 µM (0.4 µM)	10 µM (0.2 µM)	10 µM (0.4 µM)	

	Concentration (initial & final)				
	Annealing Temperature and Time (per cycle)	52°C, 30 sec	52°C, 30 sec 55°C, 30 sec 57°C, 30 sec	52°C 30 seconds	3rd

3.2 RNA Jackhammer protocol optimization using the HIVM primer panel

Initially, we attempted to multiplex the entire protocol, including reverse transcription, preliminary amplification, and the final amplification. Despite the evaluation of multiple primer concentrations and annealing temperatures, we determined that multiplexing all protocol steps produced extensive off-target formation and uneven coverage, rending the assay unsuitable for clinical use. Following this conclusion, we moved forward with a hemi-nested approach to increase specificity, while continuing to aggressively multiplex reverse transcription as well as pre-amplification steps. The final amplification step was performed in singleplex.

For each step of the process, multiple parameters were optimized towards improving yields and evenness of coverage (Table 6). When analyzing amplicon production results, factors such as presence of target amplicon and correct base pair size, band intensity, product yield, and presence (or absence) of off-target bands signaling non-specific binding, directly impacted future protocol optimization decisions. For reverse transcription, the SuperScript IV First-Strand Synthesis System improved positive amplification yields and decreased total reaction time from 150 to 60 minutes. This signified a major advancement to the original protocol, allowing completion within a single working day and increasing the feasibility for its routine clinical use. Implementing a hemi-nested PCR approach during PCR pre-amplification further improved specificity and sensitivity, with amplification rates of ~70-80% and decreased formation of non-specific products (Figure A.1).

Primer concentration was one of the most heavily optimized reaction conditions in this study, with evaluated reaction primer concentrations ranging from 0.2 µM to 10 µM. For reverse transcription and pre-amplification, primer concentrations of 0.6 µM produced the most target-

specific amplification. For the final amplification, a final primer concentration of 0.4 µM, for both the forward and reverse primers, produced the most target-specific amplification (Figure A.2).

3.3 Validation of original RNA Jackhammering protocol for HIV-1 whole genome

We then proceeded to evaluate the optimized amplicon-tiling PCR protocol using samples from the EQAPOL HIV-1 viral diversity panel (Duke Human Vaccine Institute, Durham, USA). Following sequencing with the Illumina MiSeq and data processing, this protocol produced 63.3% full genome coverage (Figure 6). Importantly, this failed to reach our targets for coverage and missed genome regions that may harbour clinically-important HIV DRMs. Therefore, we focused future work on primer sets targeting known DRMs, primarily in the *pol* region, towards the development of an assay that provides coverage across all clinically-relevant genome regions.

Table 6. List of parameters evaluated during the validation and optimization of the RNA Jackhammering protocol with the HIVM primer panel.

RT-PCR Protocol Step	Condition	Original Condition	Tested Variables	Optimized Condition	Order of Optimization
Reverse Transcription	Reverse Transcription Kit	GoScript Reverse Transcription System (Promega)	GoScript Reverse Transcription System (Promega)	SuperScript IV First-Strand Synthesis System (Invitrogen)	6th
			SuperScript IV First-Strand Synthesis System (Invitrogen)		
	Reverse Primer Concentration (initial; final)	10 µM (1.5 µM)	100 µM (10 µM)	10 µM (0.6 µM)	9th
			50 µM (5 µM)		
			25 µM (2.5 µM)		
			17 µM (1.75 µM)		
			10 µM (1.5 µM)		
			10 µM (0.6 µM)		
			Random Hexamers (50 ng/µL)		
Reverse Transcription	Annealing Temperature	50°C, 30 min & 55°C, 30 min	50°C, 30 min & 55°C, 30 min	50°C, 50 min	7th

	and Time (per cycle)		50°C, 50 min 50°C, 20 min 50°C, 10 min		
	Number of Amplification Cycles	4 (50°C) & 1 (55°C)	4 (50°C) & 1 (55°C) 1	1	8th
Reverse Transcription / Pre-amplification	Amplification Starting Point	HIV-1 viral RNA	8E5 proviral DNA	HIV-1 viral RNA	3rd
Pre-amplification / Final Amplification	PCR Kit	Platinum Taq DNA Polymerase High Fidelity (Invitrogen)	Platinum Taq DNA Polymerase High Fidelity (Invitrogen) Platinum SuperFi DNA Polymerase (Invitrogen)	Platinum Taq DNA Polymerase High Fidelity (Invitrogen)	5th
Pre-amplification	Number of Primers per Reaction	Multiplex	Multiplex Singleplex	Multiplex	1st
	Forward Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (1 µM) 10 µM (0.6 µM) 10 µM (0.4 µM)	10 µM (0.6 µM)	10th
	Reverse Primer Concentration (initial; final)	10 µM (0.4 µM)	N/A 10 µM (0.4 µM)	N/A	

	Annealing Temperature & Time (per cycle)	52°C, 30 sec	52°C, 30 sec 55°C, 30 sec 57°C, 30 sec	52°C, 30 sec	4th
	Number of Amplification Cycles	30	30 40	30	12th
	Conventional PCR modifications	N/A	Gradient PCR Nested PCR Hemi-nested PCR	Hemi-nested PCR	2nd
Final Amplification	Number of Primers per Reaction	Multiplex	Multiplex Singleplex	Singleplex	1st
	Forward Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (0.2 µM) 10 µM (0.4 µM)	10 µM (0.4 µM)	11th
	Reverse Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (0.2 µM) 10 µM (0.4 µM)	10 µM (0.4 µM)	
	Annealing Temperature & Time (per cycle)	52°C, 30 sec	52°C, 30 sec 55°C, 30 sec 57°C, 30 sec	52°C, 30 sec	4th
	Conventional PCR modifications	N/A	Gradient PCR Nested PCR Hemi-nested PCR	Hemi-nested PCR	2nd

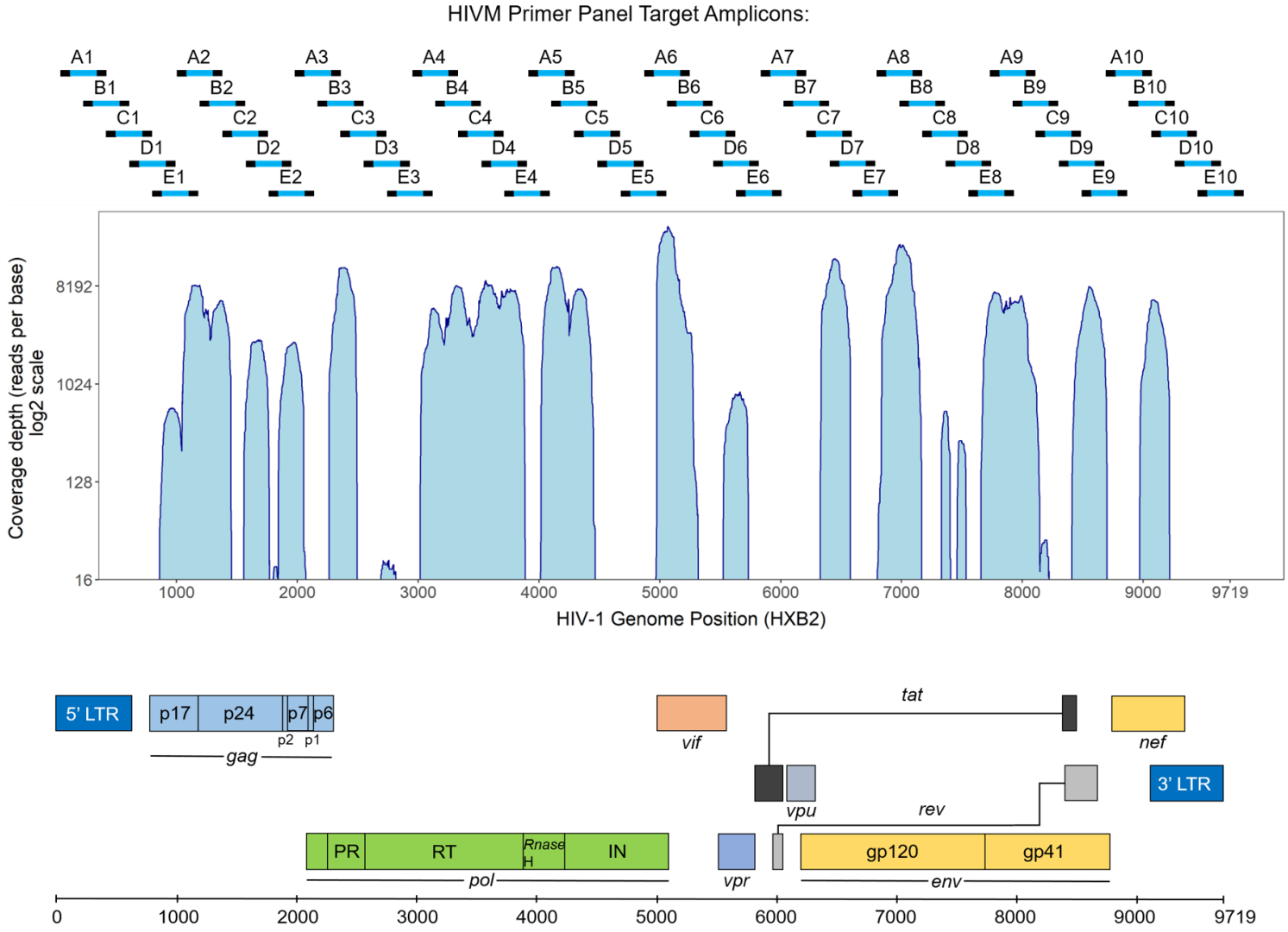

Figure 6. Coverage analysis of the optimized amplicon-tiling PCR protocol across the HIV-1 subtype B genome. Following aggressively multiplexed pre-amplification steps, singleton PCRs were performed for all 50 primer pairs. EQAPOL HIV-1 subtype B plasma was used as sample template at a viral load of 10,000 copies/mL.

3.4 Development and optimization of the PCR protocol using direct sequencing primers

We aimed to improve our protocol for high-throughput clinical use by incorporating Illumina sequencing adapters and UMIs into our primer sets. Initially, this resulted in poor amplification and the generation of non-specific products; however, through protocol optimization, primarily the introduction of bead purification steps, we were able to achieve fairly robust amplification. These protocol changes are summarized in Table 7. Specifically, we optimized the protocol in regards to primer concentration, annealing temperature, annealing time, and AMPure XP bead-to-sample ratios. This resulted in a final protocol with the following parameters: For reverse transcription, a final reverse primer pool concentration of 0.6 µM (0.15 µM each primer) and bead-to-sample volume ratio of 1.8. For pre-amplification, a final forward primer pool concentration of 0.6 µM (0.06 µM each primer), 0.4 µM for the universal reverse primer, an annealing step at 52°C for 1 minute per cycle, and a bead-to-sample ratio of 1.0. For final amplification, final concentrations of 0.4 µM for both the forward adaptor primer and the index reverse primer, and a touchdown PCR conducted at 62°C and decreased by 1°C per cycle for 10 cycles until an annealing temperature of 52°C was reached. Namely, we found the introduction of a bead-based PCR cleanup following pre-amplification, along with the use of touchdown PCR, markedly increased performance.

Table 7. List of parameters evaluated during the optimization of our HIV direct sequencing primer protocol with the HIVM primer panel.

RT-PCR Protocol Step	Condition	Original Condition	Tested Variables	Optimized Condition
Reverse Transcription	AMPure XP bead-to-sample ratios	N/A	1.8x	1.8x
			1.4x	
			1.1x	
			1.0x	
Pre-amplification	Number of primers per reaction (**first** pass of protocol)	Multiplex	Singleplex	Multiplex
			Multiplex	
	Number of primers per reaction (**second** pass of protocol; if needed)	Multiplex	Singleplex	Singleplex
			Multiplex	
	Forward Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (1 µM)	10 µM (0.6 µM)
			10 µM (0.6 µM)	
			10 µM (0.2 µM)	
	Reverse Primer Concentration (initial; final)	N/A	10 µM (0.4 µM)	10 µM (0.4 µM)
			10 µM (0.2 µM)	
			N/A	
	Annealing Temperature & Time (per cycle)	52°C, 30 sec	52°C, 30 sec	52°C, 1 min
			52°C, 1 min	
			55°C, 30 sec	
			57°C, 30 sec	

			62°C - 1°C, 1 min & 52°C, 1 min	
	Number of amplification cycles	30	30	30
			10 (62°C - 1°C) & 20 (52°C)	
Pre-amplification	AMPure XP bead-to-sample ratios	N/A	1.0x	1.0x
			0.9x	
	PCR modifications	N/A	Standard PCR	Hemi-nested PCR
			Hemi-nested PCR	
			Nested PCR	
Final Amplification	Forward Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (0.4 µM)	10 µM (0.4 µM)
			10 µM (0.2 µM)	
	Reverse Primer Concentration (initial; final)	10 µM (0.4 µM)	10 µM (0.4 µM)	10 µM (0.4 µM)
			10 µM (0.2 µM)	
	Annealing Temperature & Time (per cycle)	52°C, 30 sec	52°C, 30 sec	62°C - 1°C, 1 min & 52°C, 1 min
			52°C, 1 min	
			55°C, 30 sec	
			57°C, 30 sec	
			62°C - 1°C, 1 min & 52°C, 1 min	
			62°C - 1°C, 1 min & 48°C, 1 min	
			62°C – 0.5°C, 1 min & 48°C, 1 min	

	Number of amplification cycles	40	40	10 (62°C - 1°C) & 30 (52°C)
			10 (62°C - 1°C) & 30 (52°C)	
			28 (62°C – 0.5°C) & 17 (48°C)	
			14 (62°C – 1°C) & 30 (48°C)	
	PCR modifications	N/A	Traditional PCR	Touchdown PCR
			Touchdown PCR	

3.5 Final optimized sequencing protocol for HIV-1 drug resistance surveillance within *pol* gene

Our final sequencing protocol for HIV-1 drug resistance targets the *pol* gene, where the majority of clinically relevant drug-resistance mutations are found within the HIV-1 genome (Figure 7). Primers used within the final protocol were a combination HIVM panel primers, which we modified to allow for direct sequencing, and primers that we re-designed and implemented to replace certain HIVM primers experiencing non-amplification of the target sequence. We were able to effectively implement direct-to-sequencing primers into our protocol, which markedly increased feasibility and efficiency of the protocol by eliminating the need for costly and time-consuming NGS library preparation and the need for commercial kits. We also implemented UMIs, which allowed for the introduction of PCR bias correction steps into our in silico analyses. We outline our final sequencing protocol process in Figure 7 and Figure 8.

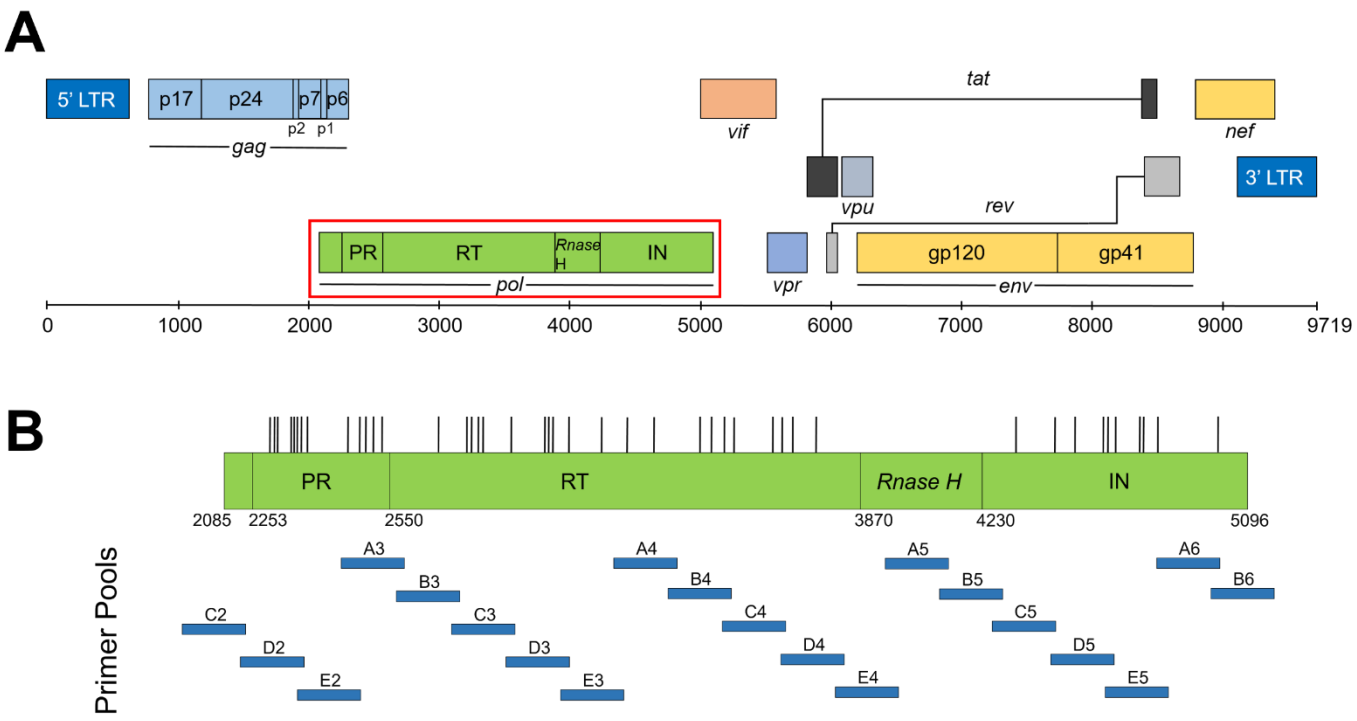

Figure 7. Target locations of direct-sequencing primers for the HIV-1 pol gene region according to the HIV-1 reference genome map. (A) HIV-1 reference genome map (HXB2) highlighting the location of the pol gene region about the entire HIV-1 genome. (B) Primer locations spanning the entire pol gene region according to the HIV-1 reference genome map (HXB2). Primers utilized are a combination of HIVM panel primers and entirely new, redesigned primers. Amino acid positions are marked for drug-resistance mutations within the protease, reverse transcriptase and integrase regions of the *pol* gene (Stanford HIV-1 drug resistance database).

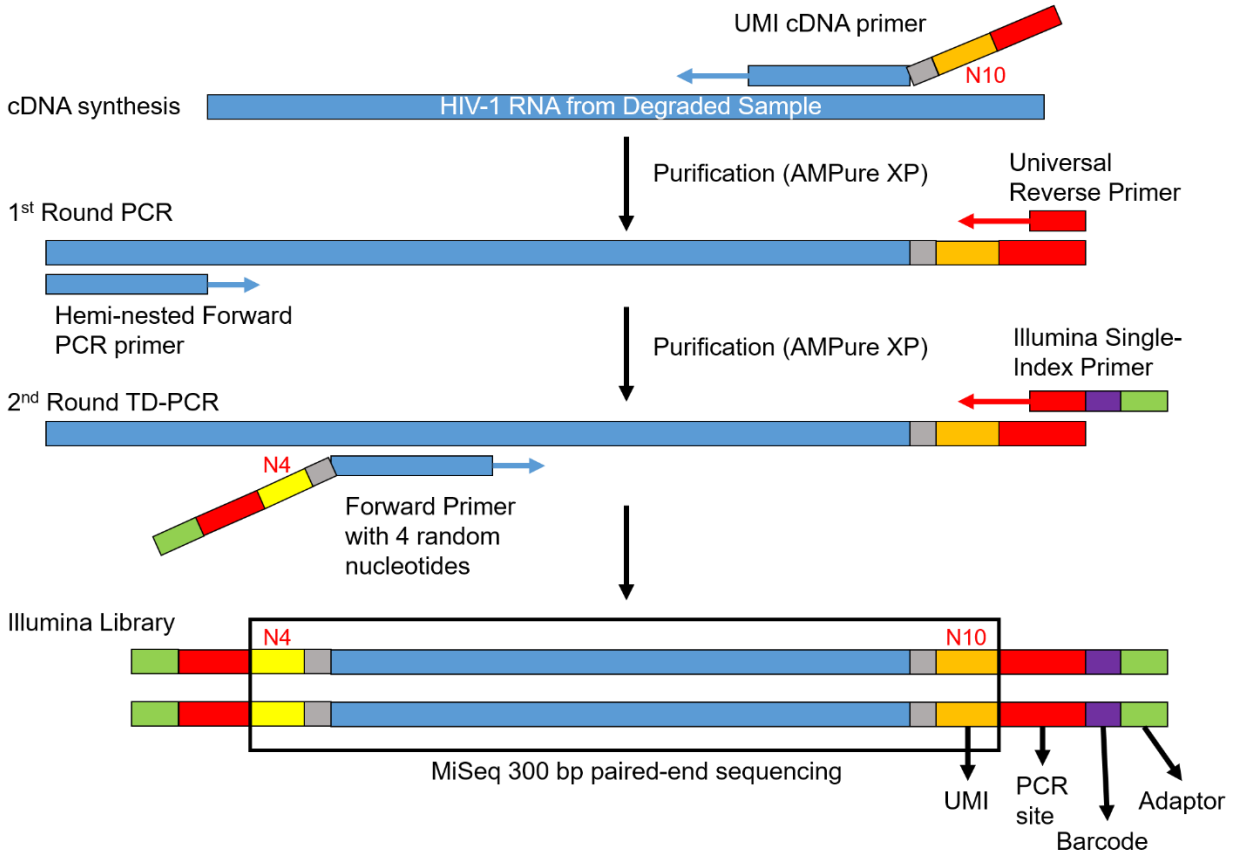

Figure 8. Direct-sequencing primer protocol including UMIs for HIV-1 drug resistance testing. Illumina adaptor (P5/P7; green) and PCR primer sites (red) are attached along with a unique molecular identifier (UMI; orange) and a spacer (GG; grey) to the reverse primer used in cDNA synthesis. The 2nd round PCR primer includes a spacer and random four-nucleotide sequence (N4; yellow) to add diversity to the samples' initial sequence reads. Illumina single-index sequences (six nucleotides; purple) are included in the 2nd round of PCR and are used for all 20 singleplex primer reactions per HIV sample.

3.6 Sequencing data analyses and UMI correction

We utilized UMIs in our protocol as a means to mitigate PCR biases common to nested and hemi-nested PCR protocols with a high number of cycles. For example, coverage plots before and after deduplication are shown in Figure 9B. There were a total number of 700,467 mapped reads (Figure 9A), decreasing to 17,736 following deduplication. The frequency of sequencing errors in barcode reads, as well as barcode swapping, was estimated by inputting artificial barcode sequences into Illumina run software that contained single nucleotide substitutions from our real experimental barcodes. Base pairs substitutions were simulated at all barcode positions for a total of six artificial barcodes sequences. For example, false reads from a change in the third base pair position of the barcode produced approximately 16 reads, as shown in Figure A.3. From this work, we estimated the error rate at each position within the barcode sequence to be approximately 1-6%, which should be considered in barcode design. As all barcodes differed at two or more base locations, we therefore expect crossover events of <0.1%, which should be taken into consideration when calling ultra-low prevalence DRMs.

The optimized protocol in full was tested on six clinical serum samples positive for HIV. The results of implementing the protocol on clinical HIV-1 subtype B samples are shown in Figure 10. Several of the samples were stored from 2013. In total, 9,510,158 reads were obtained across all samples, with 5,702,364 aligning and 58,867 surviving deduplication (Table 8). This resulted in a coverage of 99.87%. Average percent loss of NGS reads after collapsing reads from a given sample's templates was about 90%. Complete amplification of the *pol* gene was obtained in a single pass of the protocol for 4/6 of the clinical samples (66%). Complete coverage was obtained for the remaining two samples following a second pass of the protocol targeting regions of failed amplification, beginning with stored cDNA. Together, this represented a successful execution of

the optimized protocol which is functional for clinical use. In particular, this may be useful for HIVDR testing of degraded or archival clinical samples.

Template utilization (or the percent of input genomes sequenced) when using clinical specimens has been observed to be anywhere from 5-10% [87]. When determining the minimum number of raw reads needed for each sample an average number of 30 raw sequence reads per unique Primer ID are needed to ensure the majority of sequenced templates appear in the data set [87]. While the viral loads of all clinical HIV-1 samples we tested were not exactly known, they were estimated to be at least 1,000 copies/mL at the time of collection. If we assume an input RNA template number of 450 copies and 10% template utilization (i.e. 45 templates are getting sequenced by independent UMIs), and 20 amplicon regions being sequenced, we need 270,000 raw reads according to Zhou et al [87]. Taking into consideration that 40-50% of raw reads are either of low quality, or will be discarded as "offspring" sequences of independent UMIs [87], we would need 450,000 raw sequence reads per sample. In our MiSeq runs (with a 10% PhiX control to increase quality), we have observed around 10 million raw sequence reads being produced, which is concordant with what has been observed by Zhou et al [87]. With these results, we would be able to pool 22 samples for each MiSeq run and should recover enough template sequences for each region and specimen (which would be 45 template sequences per amplicon region).

Using a quick formula based on the Clopper-Pearson method (300/N), one can estimate the sampling sensitivity (in %) of detecting a variant with 95% confidence [87]. With this formula, 300 sequenced genomes are needed to have 95% confidence that a drug resistance mutation present at a true abundance of 1% is detected. Likewise, 60 sequenced templates would need to be sequenced in order to have 95% confidence that a drug resistance mutation present at a true abundance of 5% is detected.

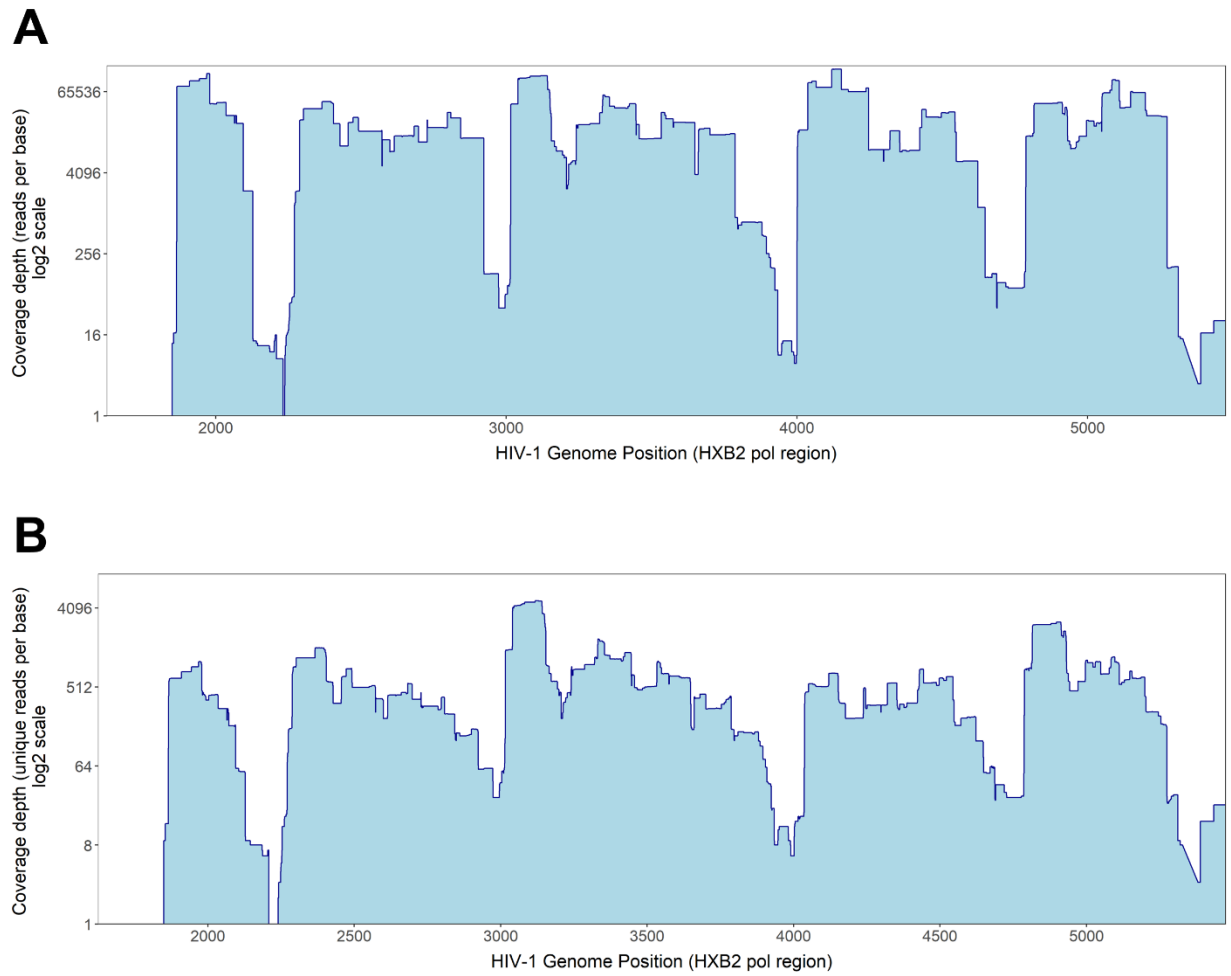

Figure 9. Coverage depth of an HIV-1 subtype B DBS sample before (A) and after (B) deduplication. Input was an HIV-1 subtype B DBS sample which had been previously contrived from an EQAPOL plasma sample designated as DEMB11US01.11. (A) All reads, including PCR duplicates, have been included in the coverage analysis of the entire *pol* gene. (B) PCR duplicates have been removed with only unique reads shown in the coverage plot analysis.

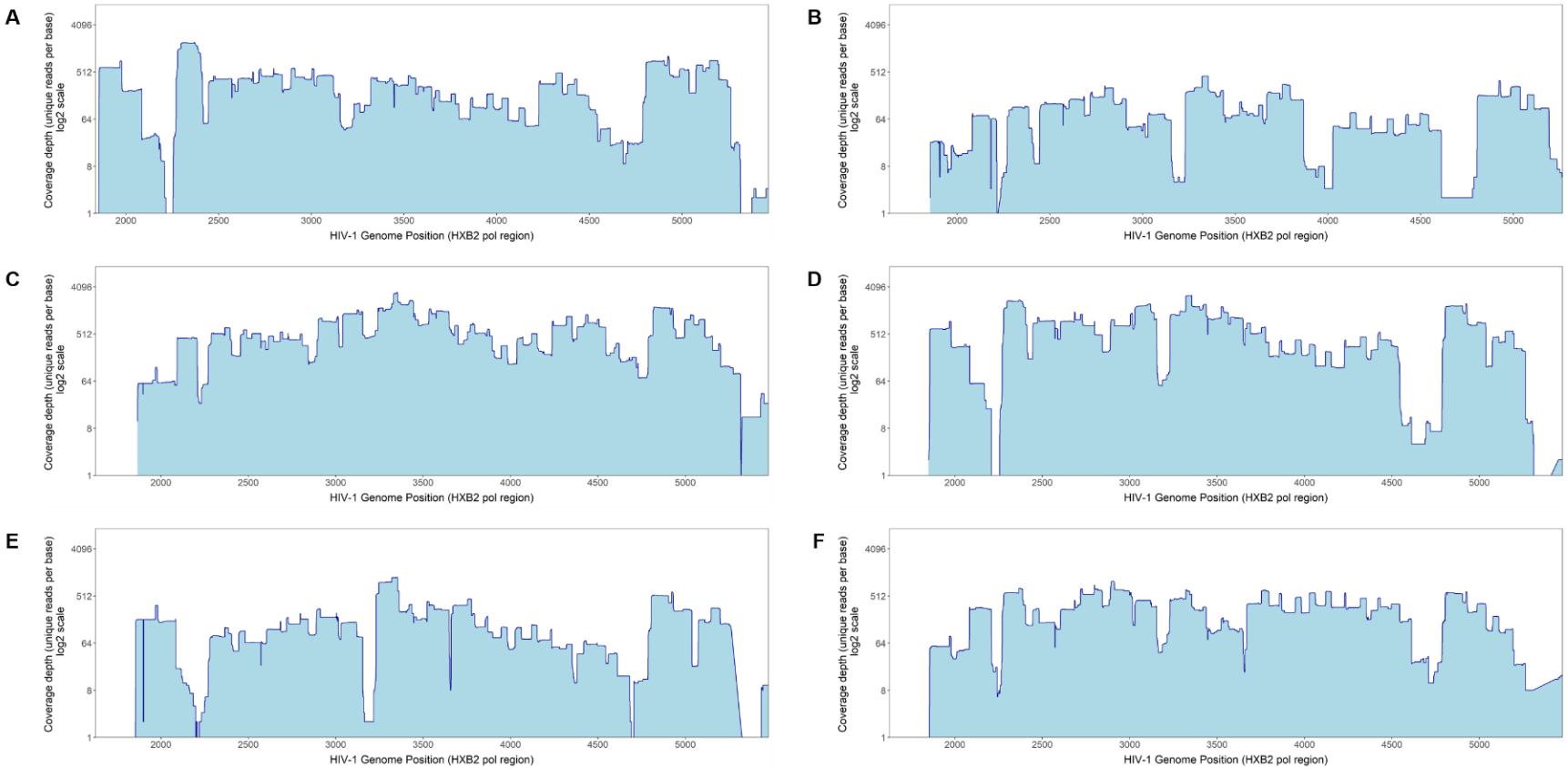

Figure 10. Coverage plot of *gag/pol* and *pol* gene of six individually archived clinical HIV-1 samples after deduplication (A-F). Sequenced samples were clinical HIV-1 subtype B serum samples archived at the JC Wilt Infectious Diseases Research Centre since 2013. Exact viral loads of each samples were not known, but were estimated at >1,000 copies/mL at the time of collection. Two samples (B and E) required a second pass through the protocol to obtain 100% coverage over all clinically-relevant DRMs. Coverage for each of these samples is reflected above after a second pass through of the protocol. The *gag* region is shown above from genome base pair positions 1879 to 2252.

Table 8. Unique read counts and depth of coverage for archived clinical HIV-1 samples

Sample Name	Raw reads per sample	Reads per sample (deduplicated)	Average depth per sample (mean)	Depth per million reads
SDR-3	1,683,245	8,391	331.47	196.92
SDR-4	1,530,095	2,551	103.21	67.45
SDR-5	1,771,120	17,347	711.65	401.81
SDR-6	1,774,020	18,061	738.57	416.33
SDR-7	1,318,254	4,489	182.152	138.18
SDR-8	1,433,424	8,028	331.09	230.98

3.7 Dried Blood Spot degradation experiment with utility of protocol with degraded samples

Ideally, an optimized amplicon tiling protocol should be compatible with DBS samples, even those that have been moderately degraded through improper shipment or storage. Therefore, we tested our protocol using progressively degraded DBS samples. To simulate poor storage conditions, samples were degraded from 1-28 days in an incubator at 37°C and 91% humidity. Any sample that initially did not achieve 100% coverage of the entire *pol* gene was reprocessed through the sequencing protocol targeting non-amplified regions. With this approach, all notable DRMs were covered by sequencing (Figure 11).

Partially degraded samples following 1-14 days of humidity and heat treatment still demonstrated a high level of coverage. Coverage achieved through this approach was 95% at day 1, 94.4% at day 7, 94% at day 14, and 99% at day 28 (Figure 12). Any comparison of day 28 to the other samples is not a direct comparison of degradation as we conducted a second-pass of the original cDNA for this sample, but not the others due to time constraints. We also compared our optimized protocol to the long-range PCR currently used for routine surveillance in the NMLB National Laboratory for HIV Genetics, a WHO-accredited, internationally-recognized, HIVDR reference lab. The comparator long-range PCR protocol amplifies ~1000 base pair amplicons covering the entire protease (PR) and reverse transcriptase (RT) regions of the *pol* gene. The long-range PCR protocols were able to amplify target regions up to 14 days of degradation, after which the template failed to amplify (Table 9).

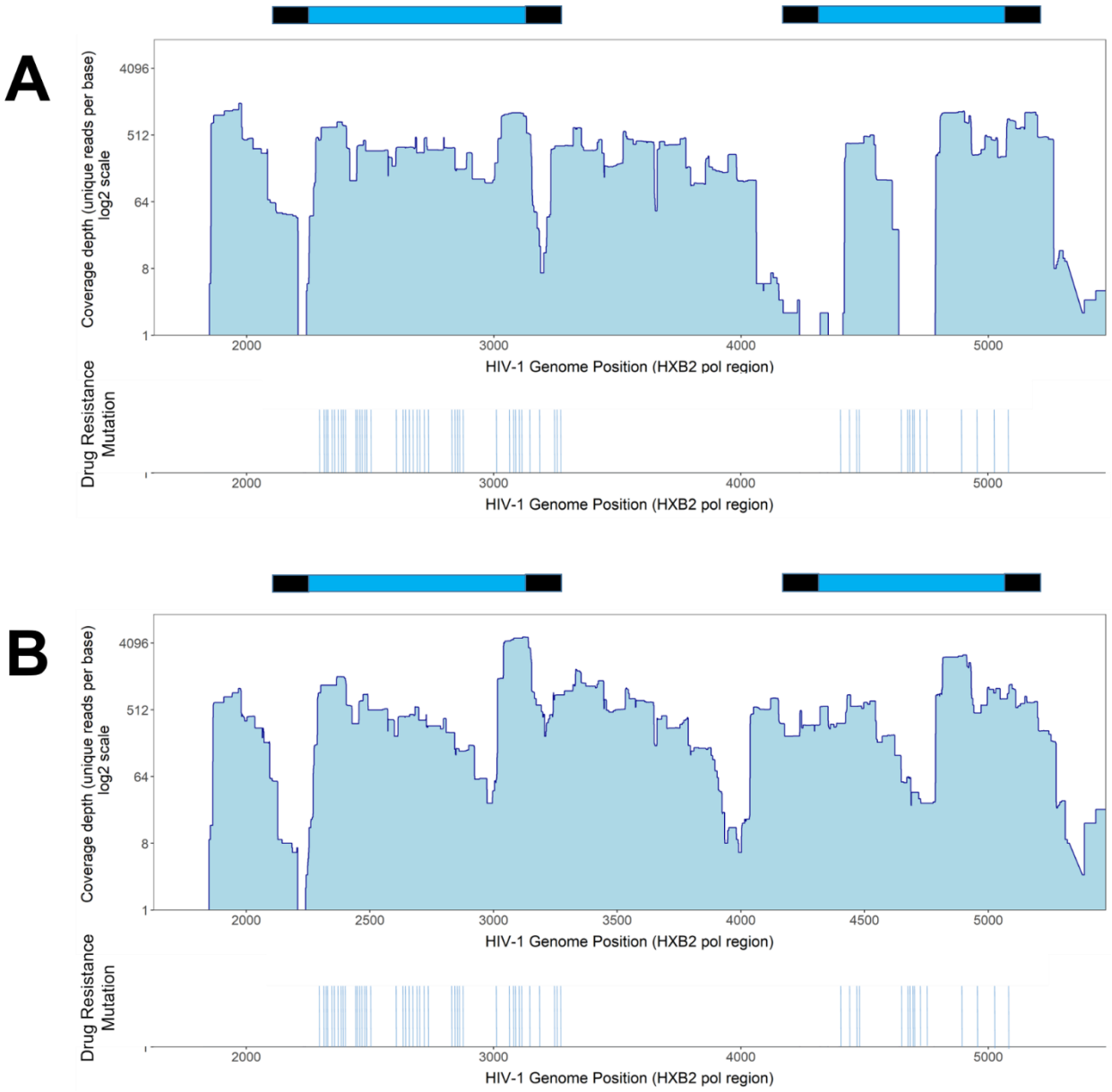

Figure 11. Coverage plot of deduplicated reads from non-degraded DBS on (A) first and (B) second pass through with our developed protocol. Known locations of common clinically-relevant DRMs are highlighted below coverage plots as indicated by vertical blue lines. Input was an HIV-1 subtype B DBS sample contrived from EQAPOL DEMB11US01.11 plasma. The NML in-house, long-range protocol amplicon locations for PRRT and IN are portrayed above the coverage plots. With our protocol, a portion of the *gag/pol* gene overlap is covered, increasing the number of potential DRMs we can capture.

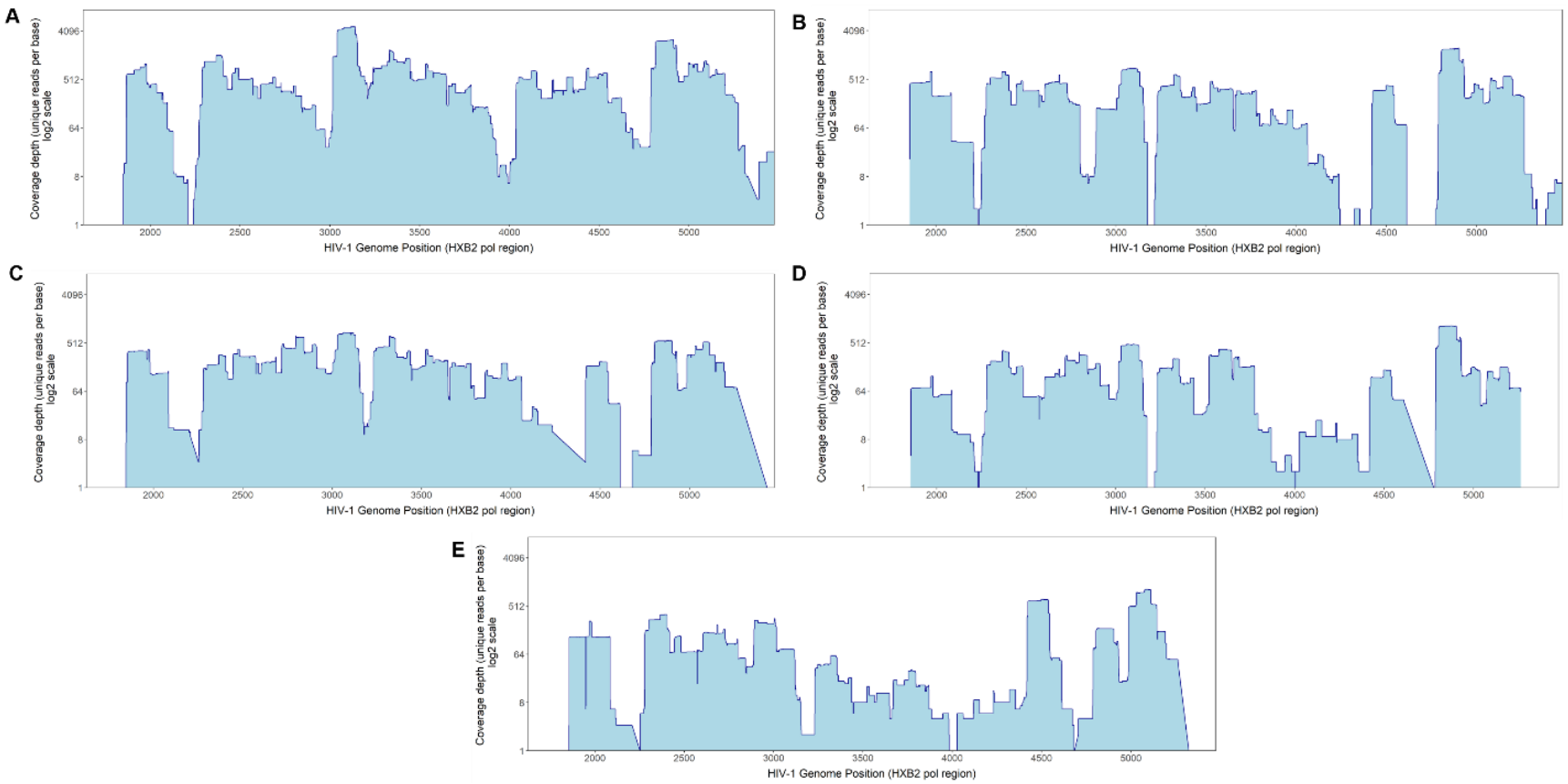

Figure 12. Coverage plots following amplicon-tiling sequencing of degraded DBS samples after deduplication. Coverage displayed is after a second pass with samples through our direct sequencing protocol. Samples were incubated at 37°C and 91% humidity and assessed at the following time points: (A) time zero, (B) 1-day, (C) 7-days, (D) 14-days, and (E) 28-days. Degradation was measured using Agilent High Sensitivity RNA Screentapes (Table 9). DBS samples were contrived in 2018 from an EQAPOL HIV-1 subtype B plasma sample dried onto DBS cards and stored at -80°C.

Table 9. Comparison of our amplicon-tiling approach to long-range PCR for amplicons of the PRRT region using degraded DBS samples.

PCR Protocol	Target Amplicon Length (bp)	Incubation Time	RNA Integrity Number (RIN^{e})[a]	Coverage of pol region
Direct-Sequencing Primers w/ UMIs	~400-500	0 days	5.1	100%
		1-day	4.5	95%
		7-days	3.2	94.4%
		14-days	2.8	94.1%
		28-days	nd	99%[b]
In-house long-range	~1000	0 days	5.1	100%
		1-day	4.5	100%
		7-days	3.2	100%
		14-days	2.8	0%
		28-days	nd	0%

[a]RIN^{e} values were measured using Agilent High Sensitivity RNA Screentapes. nd; not determined.
[b]Coverage observed at 28-days of degradation incubation is not a direct comparison to the coverage of other degradation incubation times.

Chapter 4: Discussion

According to the latest World Health Organization HIV Drug Resistance Report, 10% of adults initiating antiretroviral treatment have HIVDR to NNRTIs, with incidence rates tripled when individuals have previously used antiretroviral drugs [88]. With the prevalence of HIVDR, future research priorities include the development of innovative surveillance methods for identifying drug-resistance mutations within the HIV-1 genome [89].

Depending on the use case scenario, samples may be collected as either whole blood, plasma, serum, or as DBS. Regardless of the sample type, there is always a risk of HIV degradation during sample transport and storage [90]. In particular, highly accessible sample types, such as DBS, can be difficult to sequence – especially if improperly stored or shipped [91]. Because of this, there is a need for sequencing strategies that are compatible with degraded samples, particularly for downstream detection of HIVDR.

Sanger sequencing is still utilized heavily in clinical settings for HIVDR testing [38,92,93], with sequencing accuracy as high as 99.99% [94]. However, detection of minority HIVDR mutations (frequencies below 20%) can be challenging with traditional Sanger approaches [40]. Alternatively, NGS has the unique ability to capture these minority HIVDR mutations, which may have significant clinical implications. A number of studies have investigated the potential benefits of detecting low prevalence DRMs [95–98]. For example, Derache et al [95] examined the relationship between minority DRMs and pretreatment drug resistance in HIV-positive individuals who reported no prior antiretroviral treatment. The authors showed that they were able to detect more DRMs at a lower threshold, with 12% of patients having any observable minority resistance variant at the 5% threshold compared to 9.5% of patients having detectable DRMs at the higher 20% threshold [95]. These findings support transitioning routine HIVDR testing to NGS platforms,

as many of these DRMs at lower thresholds would be missed, due to the 20% threshold being the conventional limit for Sanger sequencing [99].

Although they can be challenging to detect clinically, low frequency DRMs can have a meaningful impact on patient response to treatment [100–102]. For example, Maruapula et al [102] demonstrated that individuals with NNRTI low-frequency mutations (<20% threshold) were 2.68 times more likely to experience virological failure compared to individuals without NNRTI low-frequency DRMs. Milne et al [101] also determined that the detection of low-frequency variants alone were associated with shorter times to virological failure, with 17 of the 30 individuals having experienced virological failure by month 8 of antiretroviral therapy. The use of NGS has been shown already to improve test sensitivity for these minority DRMs. For example, Moreno et al [103] utilized the Illumina MiSeq to identify low-frequency DRMs in individuals experiencing virological failure. Of the 22 participants identified with virological failure, the authors were able to identify minority DRMs in 11 (50%) of the individuals. Similarly, Inzuale et al [96] investigated HIVDR assessments at 1% vs. 20% thresholds. At the 1% threshold, test sensitivity to predict treatment failure at 12 months increased to 17% from 12%, albeit with a notable decrease in test specificity to 92% from 98% [96]. However, with full genome sequencing, and a building body of clinical data, HIVDR test accuracy can be expected to improve over time. As the cost of sequencing continues to decline, and through the development of user-friendly, efficient protocols for clinical drug resistance testing, it is becoming the new standard for identifying both major and minor drug resistance variants [104].

Previous protocols by Worobey et al [42] and Gryseels et al [41] have developed strategies for sequencing highly degraded HIV samples; however, these protocols are not conducive for high-throughput clinical testing. Namely, both involve extensive, time-consuming multiday procedures

and costly library preparation. Here, using the validated HIV primers of Worobey et al [42] as our foundation, we were able to develop a protocol to support HIVDR surveillance from degraded HIV-1 samples. We validated this protocol by successfully amplifying the HIV-1 subtype B *pol* gene from archived clinical HIV-1 subtype B serum samples from 2013, as well as with heavily degraded DBS samples.

The compatibility of our approach with degraded sample types is important as it allows data recovery from improperly stored or archival samples that may otherwise be unusable. The protocol can also be adapted further by expanding to other portions of the HIV genome, such as the *gag* or *env* gene regions, that also contain clinically relevant HIVDR mutations [21]. While this protocol may still be considered resource intensive for frontline testing, it provides an alternative solution in situations where routine surveillance methods fail, or for use cases where precious samples are suspected of degradation. For example, it may be of utility where archival HIV samples have been stored for several years and phylogenetic testing is desired. Additionally, this approach may prevent the need for follow up sample collection, which may be difficult or impossible, when dealing with clinical populations that are difficult to track, or international samples. Based on the data from this study, our direct sequencing protocol could be proactively used when samples have been incorrectly stored for two or more weeks, or with archival samples that have been in long-term storage.

Our initial efforts focused on amplifying the entire HIV-1 genome and examining genome coverage as an objective measure of protocol performance. We initially obtained 63% HIV-1 genome coverage, which was consistent, albeit slightly lower, than the 75% coverage observed by Worobey et al [42] as described in personal communications with authors. To cover remaining gaps, Worobey et al conducted additional passes of remnant sample through the protocol, often

with a primer walking approach. As we adapted this protocol for HIVDR surveillance, we attempted to multiplex each reaction of the entire protocol, from reverse transcription through to library preparation. Towards this objective, we utilized an exhaustive approach by evaluating various PCR techniques and parameters [105,106], including the use of hemi-nested PCR. Hemi-nested PCR (or semi-nested PCR) and nested PCR are common techniques used in HIV sequencing to achieve greater reaction specificity [107–113]. Similarly, nested or hemi-nested PCR approaches were shown to increase specificity in other amplicon-based sequencing protocols, including for HPV [114] and HCV [115].

Various other HIV sequencing protocols have achieved a near-full length genome of HIV-1, utilizing either Sanger sequencing [108,111] or NGS approaches [116–119]. Aralaguppe et al [116] developed a near full-length genome NGS protocol for HIV using clinical plasma specimens with viral loads greater than 3000 copies/mL. The protocol entailed targeting two large amplicons (approximately 5000 base pairs in length) and preparing libraries with the Nextera DNA Sample Prep Kit, with authors obtaining 100% Illumina sequencing coverage across HXB2 positions 790-9417 (*gag*-to-*nef*). Another NGS protocol developed by Ode et al [117] utilized nested RT-PCR targeting four large amplicons within *gag*-to-*nef* from clinical plasma, subsequently using the NEBNext® UltraTM DNA Library Prep Kit for Illumina. Using their protocol, they obtained >5000-fold coverage throughout the *gag* to *nef* regions of the HIV-1 genome and identified an additional 1069 minority bases in 18 samples using their deep sequencing method compared to Sanger sequencing [117]. The authors also noted differences in read mapping accuracy and coverage based on the reference sequence used, which they overcame by estimating a consensus sequence followed by iterative mapping [117]. Although both of these protocols provided excellent coverage, targeting larger amplicons likely renders them inefficient for degraded

samples, and both relied on expensive commercial library preparation kits – issues that are addressed in our current protocol.

After we had determined that utilizing this protocol for HIV-1 full-length genome sequencing was not going to be feasible, we focused our efforts on improving the protocols ability to sequence the full HIV-1 *pol* gene. By doing so, we would increase the feasibility of the protocol for clinical surveillance of HIV-1 subtype B by targeting a gene region that contains the majority of clinically relevant DRMs.

Recently, and simultaneously to this work, Hull et al [119] adapted an existing commercial amplicon-based NGS assay, Illumina COVIDSeq, for HIV-1 subtype B sequencing and drug-resistance analyses. Primal Scheme [46] was used for primer design and, similar to our protocol, primer pairs were divided into multiple primer pools. While the whole protocol was not entirely multiplexed, Hull et al [119] were able to conduct two separate highly multiplexed PCRs covering the entire HIV-1 subtype B genome. In their study, amplicons were sequenced using 150 base pair, paired-end sequencing on the Illumina MiSeq, resulting in >80% coverage for the full-length HIV-1 genome in positive control samples with viral loads as low as 488 copies/µL [119]. When identifying DRMs within *pol*, including NRTIs, NNRTIs, PIs, and INSTIs regions, analytical sensitivity of the assay on the positive control drug-resistance profile was 94%, with 64 out of the 68 DRMs correctly being identified in their dilution series [119]. However, it should be noted that all samples tested in this work were HIV-1 subtype B positive control cultures, and the authors did not evaluate performance against clinical HIV-1 plasma or serum samples, or with other subtypes. The authors acknowledge this limitation and the importance of designing future primer sets to be inclusive of the prevailing HIV-1 subtypes circulating within the testing region.

Unlike our direct sequencing protocol, the method of Hull et al necessitates extra library preparation steps after target amplification with the COVIDSeq assay. This arguably renders the protocol less suitable for routine clinical use due to its demanding workflow and higher costs. Future studies may investigate a combined approach, evaluating the primers of Hull et al [119] with the addition of UMIs and direct sequencing primers that were utilized in this study. Therefore, it may be possible to cover the entire HIV-1 genome in a manner that is feasible for high-throughput clinical testing. Further, with high DRM detection rates within *pol*, there is the potential to utilize these primers to increase coverage in the low-amplification regions identified in our study. However, this would require examination of primer compatibility characteristics such as dimer formation, melting temperature, and target site overlap before implementation would be possible.

Many HIV-1 sequencing protocols utilizing the Illumina MiSeq involve the use of traditional library preparation kits, such as Nextera XT, for the addition of sequencing components such as P5/P7 adaptors and i5/i7 barcodes [70,120]. To avoid these steps, all adapters and barcodes required for sequencing were incorporated into our primer sets. A similar approach [77] is commonly applied in 16S-based microbiome profiling [121–123]. In this study, we utilized single indices (barcodes) with our direct sequencing primers to minimize primer length. We took steps to assess the overall error rate within these indices, which ranged between 1-6%. With at least two unique bases between barcodes applied in this study, this equated to primer crosstalk of <<1%. However, if future studies are attempting to detect very low prevalence DRMs or aggressively multiplexing, the use of dual indices may be useful in eliminating cross-talk while facilitating larger pool sizes [124].

UMIs were also included in our study design as they are an innovative solution for mitigating PCR biases and identifying PCR errors. This is particularly useful in clinical approaches that rely heavily on excessive amplification, such as HIV testing, where more than 80 cycles of amplification may occur across some protocols [41,125]. However, UMIs themselves are susceptible to PCR error following excessive amplification. Recently, a study by Sun et al [126] utilized Illumina and Oxford Nanopore sequencing technologies in parallel with unique and common molecular identifiers to determine sources of error and bias. The authors concluded that PCR cycling and polymerase error rate, not sequencing read errors, were primarily responsible for error introduction into UMIs. To solve this issue, the authors designed a demultiplexing strategy that incorporated homotrimers and a set coverage approach, which exponentially reduces UMI error and outperformed standard UMI protocols [126]. With respect to our research, further examination of UMI design and error rate corrections is an important next step, especially considering the high number of PCR cycles utilized in this protocol. Applying a homotrimer-based demultiplexing method will mitigate errors in UMIs, however, this will substantially increase primer length and may affect assay performance.

Zhou et al [76] also examined the benefits of implementing UMIs (described as Primer ID in their work) into NGS workflows for HIV. By constructing their MiSeq amplicon library directly from the cDNA product through two rounds of PCR, the authors were able to accurately obtain information about initial template sampling, such as an estimate of how many input templates are converted into final consensus sequences. The authors noted that around 10% to 30% of RNA templates used in the cDNA reaction are converted to consensus sequences, with template recovery increasing as the number of raw reads increases until all available template sequences have been utilized [76]. As well, the authors determined the nucleotide substitution error rate when using the

Primer ID approach with viral RNA to be 0.01% (1 in 10,000 nucleotides sequenced), which allowed them to calculate the expected number of random errors within a reaction to define detection cutoffs for low abundance mutations [76]. With this error rate, the authors were able to determine positions with two or more mutations (>0.06% abundance) as likely to be real sequence variants within the viral sequence population [76]. With respect to our work, the implementation of UMIs allows for further examination into the sensitivity of our direct sequencing protocol and our ability to determine true sequence variants within an HIV sequence population.

In HIVDR testing, UMIs may improve the ability to accurately measure DRM frequency, which is important because low-frequency DRMs are associated with an increased risk of treatment failure [97,98]. Keys et al [127] conducted a study identifying DRMs within the HIV-1 reverse transcriptase coding domain. Using the Primer ID approach as well, they were able to reduce Illumina MiSeq errors from 24 to 1.2 errors per 10,000 nucleotides and produced less biased estimates of low-frequency DRMs [127]. Recently, Faraci et al [128] designed an HIVDR protocol for the entire HIV-1 *pol* gene using UMIs and droplet digital PCR. Using this protocol, they were able to detect 14 different minority DRMs with frequencies ranging from 3.2% to 19% for all four drug classes: PI, NRTI, NNRTI, and INSTI.

Zhou et al have extensively covered the importance of using UMIs for detection of low-frequency DRMs in HIV-1 samples [74–76,87]. One important point that the authors mention is that although protocols may claim to be able to detect 1% frequency DRMs, it does not mean they are necessarily identifying these low-frequency variants in every sample tested [87]. Utilizing UMIs to improve the detection of low-frequency DRMs, Zhou et al [74] developed an HIVDR protocol called Multiplexed Primer ID (MPID) to amplify three distinct regions (*gag*, *pol*, and *env* genes) of the HIV-1 genome, implementing UMIs during reverse transcription, as in our work.

While Zhou et al [74] consistently detected minor DRMs, they noted sampling biases for low input template samples resulted in high variance and low sensitivity. Therefore, for accurate and reliable detection of DRMs higher inputs of template input are required [40,87]. When using the Primer ID approach, Zhou et al [74] determined that having a coverage of at least 10 raw reads per Primer ID largely eliminates PCR recombinants from a typical data set. While this was not extensively examined within our work, it would be beneficial in the future to confirm that we are generating enough reads to largely eliminate PCR bias, which would further increase confidence in our protocol to accurately detect DRMs. Zhou et al [74] also noted poor generation of consensus sequences when attempting to multiplex steps of their protocol. For example, when all three primers targeting the *pol* gene (PR, RT, and IN) were multiplexed, a 20-40% reduction in amplification was observed [74]. Likewise, in our protocol we identified a similar trade off when multiplexing our protocol, with uneven coverage within close-proximity genome regions likely the result of competitive interactions during amplification.

To demonstrate feasibility, we validated our finalized protocol on degraded DBS samples and directly compared its performance for long-range PCR approaches. For Northern, remote and isolated communities, plasma and serum samples may not always be the most accessible sample type to use for HIVDR. As a result, DBS have become increasingly popular in resource-limited settings, such as Northern, remote, and isolated communities [33]. In Canada, the NML houses a WHO-accredited HIVDR reference lab with strong international collaborations, with DBS samples often shipped from areas of Africa with high HIV prevalence rates. Unfortunately, samples received frequently arrive in a degraded state due to long transportation times and sub-optimal storage conditions. In our study, simulated degradation of DBS samples and assessed our

coverage across HIV-1 *pol*. With our direct sequencing protocol, we were able to obtain near full coverage of the *pol* gene in all heavily degraded samples when primary testing failed.

Consistent with this work, previous studies have shown improper sample storage leads to poor amplification in long-range PCR. For example, Garcia-Lerma et al [31] were unable to perform HIV-1 genotyping on DBS samples stored beyond two weeks at 37°C and 91% humidity, in agreement with the recommended short-term storage guidelines from the World Health Organization [34]. The authors did note that only one 50 µL DBS was extracted per sample instead of the conventional two spots per sample, which may have improved amplification rates.

DBS samples provide a small volume of blood and thus a limited amount of template for testing. By multiplexing both our reverse transcription and pre-amplification steps we facilitate DBS testing by reducing the amount of template required. In addition, this preserves enough genetic material so that a second pass through of the protocol can be conducted if needed, beginning with reverse transcribed cDNA, facilitating incredibly high coverage (up to 100%) even in degraded DBS samples.

Probe capture presents another approach to HIV-1 genome sequencing and HIVDR genotyping, reviewed extensively by Munyuza et al [67]. Using probe capture enrichment methods for HIV, Yamaguchi et al [69] achieved full genome coverage at 10-100 fold lower input concentrations when applying probe capture to their amplicon tiling protocol [69,70]. Likewise, Miyazato et al [129] used target enrichment for next-generation sequencing of proviral DNA. The authors found that 99.5% of reads mapped back to the proviral genome with libraries prepared using enrichment methods, compared to just 1.9% when prepared without enrichment [129]. Together, the above studies demonstrate that implementing target-enrichment methods can greatly increase on-target reads; however, there are notable drawbacks to the technique. Library

preparation steps are still time consuming and costly, which much be performed ahead of probe capture [66], as well as the development and optimization of effective probes sets can be challenging. In addition, it possesses many of the same challenges as amplicon tiling, including uneven coverage between enrichment sites and complex, multilayered protocols.

Oxford nanopore sequencing is another promising solution for HIV-1 genome sequencing, for both HIVDR testing and phylogenetics. As a low-cost and low barrier solution [130], nanopore sequencing has the potential to generate large amounts of data and incredibly long reads. While this technology can be applied to total extracted DNA/RNA without amplification [130], a low number of on-target reads would be expected. A possible solution for this issue is adaptive sampling, a process that allows for the selective expulsion of off-target reads from the sequencing porin. A recent study by Martin et al [68] showed up to a 5-fold increases in on-target reads of low-abundance organisms using this approach; however, they determined approximately 40% of on-target reads were being falsely rejected by the GridION [68]. As an additional limitation, nanopore sequencing has historically shown lower accuracy as compared to Illumina; however there have been recent and significant advances to the technology. Minei et al [131] demonstrated that the MinION boasts an accuracy of 87.4%, as demonstrated by compared to previous accuracies of 64.3% for template reads [132]. The latest flow cell released by Nanopore, R10.4.1, with the ligation sequencing kit V14 now boasts a raw read accuracy of 99% with a Q20 score [133], demonstrating the rapid evolution of the technology. Although nanopore sequencing is still limited by accuracy and sequencing depth, the notable progress on these fronts, along with its inherent benefits including long reads, a simplified workflow, and affordability, make it an excellent future candidate for HIV whole genome sequencing.

With increasing success of nanopore sequencing for viruses such as Ebola [134], Zika [46], influenza [135], and SARS-CoV-2 [47,136], other studies have evaluated nanopore sequencing for HIV genotyping [137–139]. Mori et al [137] used nanopore sequencing to examine recombinant forms of HIV-1, successfully identifying recombinant forms and assigning subtype information. The authors did note two major limitations with using the technique. First, sequencing errors can occur in homopolymer or tandem repeat regions, which are frequent within the HIV-1 genome, especially within the *pol* gene [137]. Second, any minority reads obtained by nanopore sequencing are masked by major sequences during consensus sequence construction, leading to sequence diversities that did not reflect the source and inaccurate measurement of HIV-1 minority variants [137]. Ng et al [138] also examined the use of nanopore sequencing, using the approach for the identification of mixed HIV quasi-species from clinical plasma samples collected from 2002 to 2014. The authors first used target amplification of the 1413-7363 base pair genome region of HIV-1 (reference NC_001802.1), before PCR cleanup and sequencing with either nanopore-, Illumina-, or Sanger-based approaches. The researchers noted the precision and recall of the Oxford nanopore was 93.8% and 99.2% respectively, as compared to the other technologies [138]. With these findings, we could potentially utilize Nanopore instead of Illumina after our amplicon-tiling protocol. This could allow us to shorten our primers by eliminating the need for Illumina adapters or sequencing sites. This option becomes more feasible as Oxford nanopore chemistries continue to improve.

There are several potential advantages of our newly developed protocol. One advantage is our ability to cover multiple genes, both the *pol* gene and the *gag-pol* gene overlap, within a single experiment to obtain more sequence information, including DRMs, from a given HIV-1 viral RNA sample. Another advantage is the utilization of UMIs to enable accountable quantification of all

DRMs that may be present within the *pol* or *gag-pol* gene regions, including LADRVs. Lastly, our design and development of direct sequencing primers allows us to bypass using the Nextera XT DNA kit for library preparation, which comprises two-thirds of the reagent cost of regular NGS HIVDR assays [26].

While there are several potential advantages, there are limitations with our proposed method. According to the IAS-USA [21] and the Stanford University HIV Drug Resistance Database [19], the majority of clinically relevant DRMs are found within the HIV-1 *pol* gene region. However, DRMs found within the *gag* [22] and *env* [23] regions can also play a role in drug-resistance. With our direct sequencing protocol, there is the potential to identify DRMs within the *gag-pol* and *pol* gene regions that may be interact to confer drug-resistance. However, due to our limited range across the genome, we may fail to capture some atypical HIV mutations that may, through cross genome interactions, influence resistance profiles [140,141]. In addition, the protocol design is fairly complex compared to conventional long-range PCR requiring multiple bead purification steps, management of multiple primer pools, and individual singleplex final amplification reactions. There may also be a need to reprocess cDNA when working with low quality or low input samples to obtain 100% coverage across clinically relevant DRMs. Taken together, this protocol in its current state is likely not be suitable to replace high-throughput long-range PCR protocols for routine testing [138]. However, the proposed method could be utilized as an adjunctive protocol for samples that fail to amplify during first line testing, or for those that are suspected of having undergone heavy degradation.

Conclusion and Future Directions:

In this study, we have presented a comprehensive protocol for HIVDR testing, which has been extensively optimized and streamlined for clinical use. The steps outlined in our protocol are

highly adaptable for common laboratory robotics, and were already largely automated in this study on the Eppendorf epMotion 5075 (Hamburg, Germany). This developed protocol covers the majority of clinically relevant DRMs; however, future work could expand the target sites to cover other genome regions or potentially the entire genome. For example, the protocol can be extended and further adapted to cover the upstream capsid gene targeted by a new category of ART drugs. Enhancing the protocol in this way would facilitate more robust phylogenetic analyses from sequencing data, and future-proof the assay against novel or emerging DRMs. Although this protocol may be too cumbersome for routine HIVDR surveillance, it provides utility for HIVDR testing and phylogenetics when samples are difficult to amplify.

We were able to validate our approach with the most commonly used sample types for HIV testing, including plasma, serum, and DBS. This includes compatibility studies on degraded DBS samples, and comparison to long-range PCR. However, the subtypes evaluated in this study were limited to only HIV-1 subtype B. Due to mutations, insertions, deletions, and other genetic differences between HIV-1 subtypes, there are still questions about how the protocol will perform with other subtypes, such as A,C, or D. It is an important next step to challenge this assay with a diverse panel of HIV clinical specimens to screen for potential primer dropouts. In this instances, it may be important to revisit the design of specific primer sets, potentially incorporating additional degenerate bases to accommodate a broad range of subtypes.

The development of this protocol contributes novel findings to the ever growing body of NGS-based HIVDR assays. Among these findings include the ability to design and implement direct sequencing primers (including UMIs) to sequence the entire HIV-1 subtype B *pol* gene, including overlapping regions, in archival clinical samples. In doing so, further studies focused towards phylogenetics, HIVDR contact-tracing, and transmission of DRMs can potentially utilize

this protocol for further analysis of HIVDR within a given population. As well, our protocol has utility for processing degraded sample types, such as DBS samples, while still being able to identify DRMs found within the sample. This finding could potentially improve the diagnostics landscape and accessibility for HIVDR testing. It could be particularly useful for molecular epidemiology of HIV circulating within remote Canadian communities or internationally, or in situations where samples received by the NML from resource-limited settings are difficult to amplify. Ultimately, this serves as a critical addition to the diagnostic toolkit that ensures HIV clinical testing for drug resistance is as accurate and informative as possible, even under challenging circumstances.

Appendix

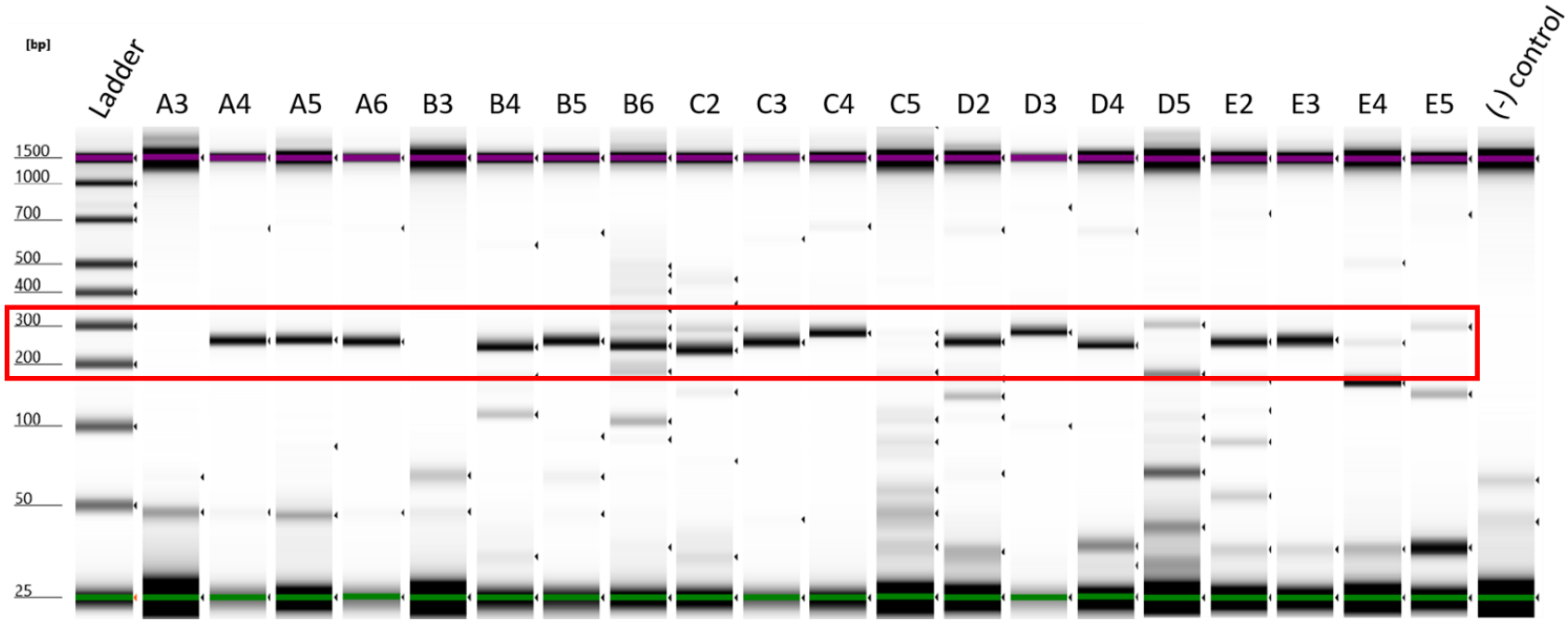

Figure A.1. Increased target amplicon yield and specificity using SuperScript IV with a hemi-nested approach. Positive amplification rates of 75-80% are shown targeting the entire *pol* region. Target amplicons were found within the 200-300 bp range (red). Primer names are shown above each respective lane. Gel image was generated using the Agilent TapeStation 4200 and D1000 screentapes.

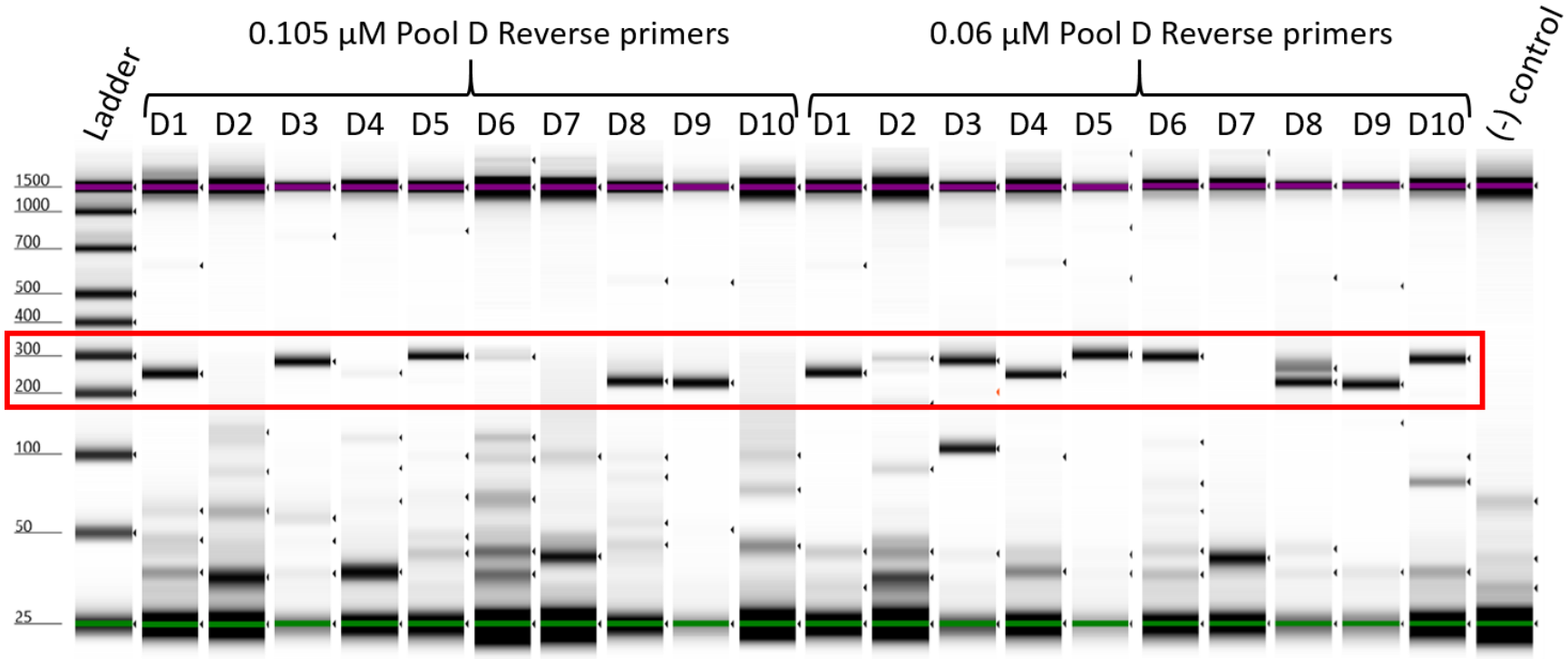

Figure A.2. Determination of optimal reverse primer concentration for reverse transcription. Individual reverse primer concentrations of 0.06µM within a multiplex reaction showed the highest number of target amplicon generation compared to other concentrations, such as 0.105 µM per primer. Target amplicons were found within the 200-300 bp range (red). Primer names are shown above each respective lane. Gel image was generated using the Agilent TapeStation 4200 and D1000 screentapes.

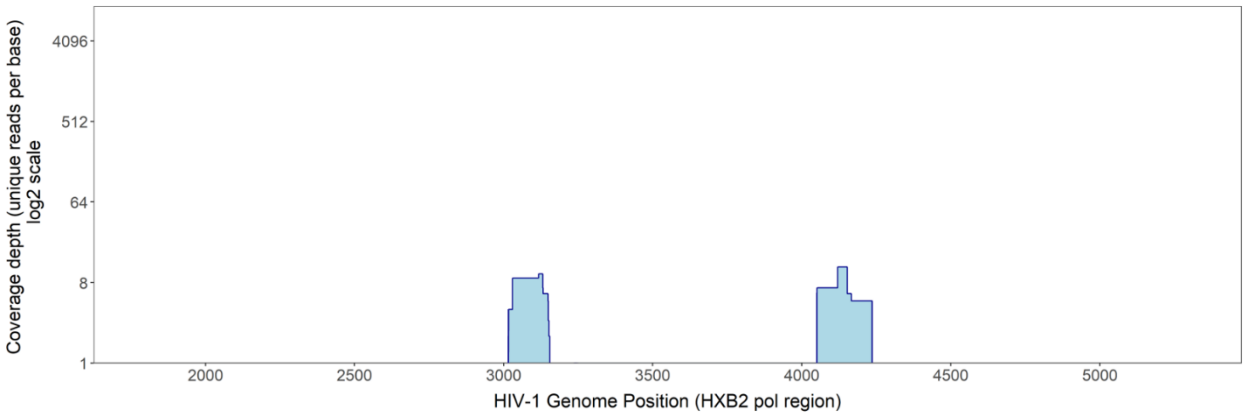

Figure A.3. Mapping deduplicated reads with a false Illumina barcode with the third barcode base swapped. Reads mapping to the HXB2 region would have been added to the original barcode but due to slight errors with Illumina sequencing, the barcode was swapped and caused a small amount of reads to be grouped into a non-existing barcode.

Table A.1. Original HIVM panel primers from Worobey et al.

Forward Primer	Sequence (5' → 3')	Target / Gene
HIVM1F	CCTCAGACCMTTTTAGTCA	5' LTR-*gag*
HIVM2F	GYTAAGGCCAGGRGGAA	*gag*
HIVM3F	RCCCTCTATTGTGTRCATC	*gag*
HIVM4F	ARGCMATATCACCTAGAAC	*gag*
HIVM5F	TCAATGARGAAGCTGCA	*gag*
HIVM6F	CACCTATCCCAGTAGGAGA	*gag*
HIVM7F	GGATGACAGAAACCTTGTTG	*gag*
HIVM8F	TGTCAGGGAGTRGGRGGA	*pol*
HIVM9F	WTGYACTGAGAGACAGGCTA	*pol*
HIVM10F	CAGRTCACTCTTTGGCAA	*pol*
HIVM11F	ATAGGGGGAATTGGAGGKT	*pol*
HIVM12F	AAAGCCAGGAATGGATG	*pol*
HIVM13F	TGGGCCTGAAAAYCCAT	*pol*
HIVM14F	GTAYTGGATGTRGGTGATG	*pol*
HIVM15F	AGGATCACCAGCAATATTCC	*pol*
HIVM16F	TCAGAARGAACCYCCATTCC	*pol*
HIVM17F	CCTTAGRGGARCCAAAGCA	*pol*
HIVM18F	GGAGTGTATTATGACCCATC	*pol*
HIVM19F	TTAACAGAGGCAGTGCA	*pol*
HIVM20F	GARAAAGAACCCATARTAGG	*pol*
HIVM21F	CAGGATTCRGGAYTAGAAG	*pol*
HIVM22F	RCTGGAATCAGGAAAGTAC	*pol*
HIVM23F	YCCAGGAATATGGCAAYTAG	*pol*
HIVM24F	GCAGGAAGATGGCCAGT	*pol*
HIVM25F	AAAGAAAAGGGGGGATTGG	*pol*
HIVM26F	AGGGGCAGTAGTAATACAAG	*pol-vif*
HIVM27F	GCAGGTGATGATWGTGTG	*vif*
HIVM28F	TAGRGGATGCTARATTGGT	*vif*
HIVM29F	AGTCKCCRTAGAATGGAGGA	*vif*
HIVM30F	AAAGCCRCCTTTGCCTA	*vif-vpr*
HIVM31F	GAGGAGCTTAAGARTGAAGC	*vpr*
HIVM32F	TGYCRACATAGCAGAATAGG	*vpr*

HIVM33F	SCTTAGGCATCTCCYATGG	*vpu*
HIVM34F	GCAATAGTTGTGTGGWCYA	*env*
HIVM35F	AAWTGTGGGTCACAGTC	*env*
HIVM36F	RAYATGGTAGAACAGATGC	*env*
HIVM37F	YTAACCCCACTCTGTGT	*env*
HIVM38F	MTTACACARGCCTGTCCA	*env*
HIVM39F	CASCACAGTACAATGTACAC	*env*
HIVM40F	GTAYAAGACCCAACAAC	*env*
HIVM41F	GGACCCAGAARTTGTAAYG	*env*
HIVM42F	ACATGTGGCAGAAAGTAGG	*env*
HIVM43F	GGGACAATTGGAGAAGTGA	*env*
HIVM44F	GACGGTACAGRCCAGA	*env*
HIVM45F	AGARTCCTGGCTGTRGAA	*env*
HIVM46F	ATCGCARAACCARCAAG	*env*
HIVM47F	GACAGGCCCGAAGGAA	*env*
HIVM48F	GAGGAYTGTGGAAMTTCTG	*env*
HIVM49F	GACAGGGCYTRGAAAGG	*nef*
HIVM50F	TGGCTAGAAGCACAAGAG	*nef*-3’ LTR
Reverse Primer	**Sequence (5’ → 3’)**	**Target / Gene**
HIVM1R	YTCCCTGCTTGCCCAT	5’ LTR-*gag*
HIVM2R	CTAARGCTTCCTTGGTGTC	*gag*
HIVM3R	CTGAAAGCCTTYTCTTCTAC	*gag*
HIVM4R	AAYAGGCCCTGCATGCA	*gag*
HIVM5R	GTTCYTTTGGTCCTTGTYT	*gag*
HIVM6R	TCTGGGTTYGCATTTTGG	*gag*
HIVM7R	TGYCCTTCTTTGCCACA	*gag*
HIVM8R	AAGGCCAGATYTTCCCTA	*pol*
HIVM9R	CYTTTAGTTGCCCYCCTA	*pol*
HIVM10R	TTGACAGGTGTAGGTCCT	*pol*
HIVM11R	CTTCTGTYAATGGCCAT	*pol*
HIVM12R	TGAACTTCCCAGAARYC	*pol*
HIVM13R	GCAGTATACTTYCTRAAGTC	*pol*
HIVM14R	TGYYGCCCTATTTCTAAGTC	*pol*
HIVM15R	GGATGGAGTTCATAACCCAT	*pol*

HIVM16R	CYAGTTCTAGCTCTGCTTC	*pol*
HIVM17R	CACCCCTCRTYCTTGCAT	*pol*
HIVM18R	TCCMCCATGYTTCCCATG	*pol*
HIVM19R	CTRTTRGCTGCCCCATC	*pol*
HIVM20R	TKATCTGGYTGTGCTTG	*pol*
HIVM21R	TCTTGGGCCTTATCTATYCC	*pol*
HIVM22R	GCTACCAGRATAAYTTTTCC	*pol*
HIVM23R	GATYCCYGCCCACCAA	*pol*
HIVM24R	TTCCCCTGCACTGTAYC	*pol*
HIVM25R	ACCTGCCRTCTGTTTTCC	*pol*
HIVM26R	GTGGGAYRTGTACTTCTGA	*pol-vif*
HIVM27R	RCCCAAATGCCAGTCYCT	*vif*
HIVM28R	CCTAGGACTAACTMTAYGTC	*vif*
HIVM29R	CTTGTTCCATCTATCCTCTG	*vif*
HIVM30R	TTATGGCYTCCACTCCT	*vif-vpr*
HIVM31R	GGCTCTARTYTAGGATCTAC	*vpr*
HIVM32R	MTCTKCGTCGCTGTCTC	*vpr*
HIVM33R	TGCCACTGTCTTCTGCT	*vpu*
HIVM34R	CTTCYTTCCACACAGGTAC	*env*
HIVM35R	CACATGGCTTTARGCTT	*env*
HIVM36R	CTGAGGTRTTACAAYTTATC	*env*
HIVM37R	GTATGGGAATTGGCTSA	*env*
HIVM38R	CAGATTYRTTCAGCTGTAC	*env*
HIVM39R	ATGCTCTCCCTGGTCCT	*env*
HIVM40R	YTTCTGGGTCCCCYCCT	*env*
HIVM41R	CTRATGGGAGGRGCAT	*env*
HIVM42R	TYTGCACCACTCTTCTCT	*env*
HIVM43R	TGTTGCGCYTCAATAGC	*env*
HIVM44R	CCAAGGCACAGYAGTG	*env*
HIVM45R	TCCACAAACTTGCCCAT	*env*
HIVM46R	CTCTCTCTCCACCTTCTYC	*env*
HIVM47R	CTRTCTGTCCCCTCAGYT	*env*
HIVM48R	ACCAYTTGCCACCCAT	*env*
HIVM49R	TGGTCTTAAAGGWACCTGRG	*nef*
HIVM50R	GWAGCACCAYCCAAAGGT	*nef*-3’ LTR

Table A.2. Original HIVL panel primers from Worobey et al.

Forward Primer	**Sequence (5' → 3')**	**Target / Gene**
HIVL1F	TCACTCCCARMRAAGACA	5' LTR
HIVL2F	CCAGGGATCAGRTWYCCA	*gag*
HIVL3F	GAGGCCAAMDAAGGAGAGA	*gag*
HIVL4F	TTTGACTAGCGGAGGCTA	*gag*
HIVL5F	CTTTCAGCCCAGAAGTRA	*gag*
HIVL6F	GGACATAARACAAGGRCCAA	*gag*
HIVL7F	CCCACCAGMRGARAGCTT	*pol*
HIVL8F	CCTATTGARACTGTACCAGT	*pol*
HIVL9F	GACTTAGAAATAGGGCRRCA	*pol*
HIVL10F	GARCCAGTACRTGGRGTGT	*pol*
HIVL11F	GCAGCYAAYAGGGARACT	*pol*
HIVL12F	GTAGTCCAGGAATATGGCAA	*pol*
HIVL13F	GAAAGGACCAGCAAARCT	*pol-vif*
HIVL14F	TACTTGGCACTARYARCA	*vif-vpr*
HIVL15F	GCATCTCCTATGGCAGGA	*vpu-env*
HIVL16F	CWGTRTGGAAAGAAGCAA	*env*
HIVL17F	CACCACAAGCRTRRGAGAT	*env*
HIVL18F	GTCTAGCAGWAGARGAGRTA	*env*
HIVL19F	AATGTATGCYCCTCCCAT	*env*
HIVL20F	AAGARTCCTRGCTGTGGA	*env*
HIVL21F	MGAAGRAGAMGGTGGAGAGA	*env*
HIVL22F	CKCTTGAGAGACTTACTMT	*env*
HIVL23F	AGTAGCTGRGGGGACAGA	*env-nef*
HIVL24F	GGDTGGTGCTWCAAGCTAG	*nef*
HIVL25F	TYKCWACAAGGGACTTTCC	3' LTR
Reverse Primer	**Sequence (5' → 3')**	**Target / Gene**
HIVL1R	CTGGYTYTACTTTCGCTT	5' LTR
HIVL2R	TTTGGCGTACTCACCAGT	*gag*
HIVL3R	GCACCCATCTCTCTCCTT	*gag*
HIVL4R	CATTCTGCAGCYTCCTCA	*gag*
HIVL5R	TCCTCCYACTCCCTGRCA	*gag*
HIVL6R	CCCYCCTAYCTTTATTGTGA	*gag*

HIVL7R	CTTCCCAGAARTCTTGAGT	*pol*
HIVL8R	ARATGYTGTCTCAGYTCCT	*pol*
HIVL9R	RTCCMCCATGYTTCCCAT	*pol*
HIVL10R	TCTRATCCYGAATCYTGCA	*pol*
HIVL11R	GCTACATGRACTGCTACYA	*pol*
HIVL12R	CCTGCCATCTGTTTTCCA	*pol*
HIVL13R	GCTTGTTCCATCTATCYTCT	*pol-vif*
HIVL14R	GAGAARCYTGATGAGTCTBA	*vif-vpr*
HIVL15R	RCTRACACAGAGTGGGGTT	*vpu-env*
HIVL16R	CTGGYCTAATTCCATGTGT	*env*
HIVL17R	CCCYCCTGAGGAKTGMTT	*env*
HIVL18R	CTTCTCCAATTGTCCMTCA	*env*
HIVL19R	GYAGTGGTGCARATGAGT	*env*
HIVL20R	CCAGAAKTTCCACARTCCT	*env*
HIVL21R	CYYCYTCCTCTTGTGCTT	*env*
HIVL22R	GTRAAYTAGCCCWTCCAGT	*env*
HIVL23R	CTCTCCTTYATTGGCCTYT	*env-nef*
HIVL24R	CAGCTGCTTATATGCAGSA	*nef*
HIVL25R	CACACTGACTAAAAKGGTCT	3’ LTR

Table A.3. Original HIVR panel primers from Worobey et al.

Forward Primer	Sequence (5' → 3')	Target / Gene
HIVR1F	CAGAGGAGMTCTCTCGA	*gag*
HIVR2F	GTGCGAGAGCGTCRGT	*gag*
HIVR3F	TCAGACAGGATCAGAAGARC	*gag*
HIVR4F	MARGTCAGCCAAAATTACC	*gag*
HIVR5F	CACCATGYTAAACACAGTG	*gag*
HIVR6F	CCARATGAGAGAACCAAG	*gag*
HIVR7F	ACAAGGACCAAARGAACC	*gag*
HIVR8F	ACCAGCRGCYACACTAG	*gag*
HIVR9F	AATTGCARGGCCCCTAG	*gag-pol*
HIVR10F	ACARCTCCCYCTCAGAAG	*pol*
HIVR11F	CAGGAGCAGATGATACAGT	*pol*
HIVR12F	YCCTATTGARACTGTACCA	*pol*
HIVR13F	RAACTCAAGAYTTCTGGGAA	*pol*
HIVR14F	AYAATGAGACACCAGGGAT	*pol*
HIVR15F	GAAAYAGGGCRGCATAGAAC	*pol*
HIVR16F	CCAGAAAARGACAGCTGGA	*pol*
HIVR17F	GAACTGGCAGAAAACRGRG	*pol*
HIVR18F	ACTACCYATACAAAARG	*pol*
HIVR19F	GTAGGAGCAGAAACYTT	*pol*
HIVR20F	CAGAAGACTGARTTACAAGC	*pol*
HIVR21F	CAGCACACAAAGGAATTGG	*pol*
HIVR22F	CCAGCTGTGATAAATGYCAG	*pol*
HIVR23F	AGGRCAGGAAACAGCA	*pol*
HIVR24F	CCAAAGTCAAGGAGTAGTAG	*pol*
HIVR25F	CAGGGGAAAGAATARTAGAC	*pol*
HIVR26F	GTGCCAAGAAGAAAAGC	*pol*
HIVR27F	GGATGAGGATTARMACATGG	*vif*
HIVR28F	GTCTCCRTAGAATGGAGGAA	*vif*
HIVR29F	ATGGAACAAGCCCCAGA	*vpr*
HIVR30F	TCAGAATTGGGTGTCGA	*vpr-vpu*
HIVR31F	AGGAAGAAGCGGAGACA	*vpu-env*
HIVR32F	GAAARAGCAGAAGACAGTG	*env*

HIVR33F	CACACATGCCTGTGTACC	*env*
HIVR34F	TGGGATCAAAGCCTAAAGC	*env*
HIVR35F	GTCATTACACARGCCTGT	*env*
HIVR36F	GGACCATGTACAAMTGTCAG	*env*
HIVR37F	CACRRACAATGCTAAAACC	*env*
HIVR38F	GGAGGGGAATTTTTCTACTG	*env*
HIVR39F	CAGGGCTGMTATTAACAAG	*env*
HIVR40F	CAGCAGGAAGCACYATGG	*env*
HIVR41F	CAACAGCTCCTRGGGAT	*env*
HIVR42F	GGCAAGTTTGTGGAATTGG	*env*
HIVR43F	AGGCAGGGATAYTCACCAT	*env*
HIVR44F	GATCTGMGGARCCTGT	*env*
HIVR45F	GAGGGGACAGATAGGRTT	*nef*
HIVR46F	GTGGGAGCAGYATCTC	*nef*-3’ LTR
Reverse Primer	**Sequence (5’ → 3’)**	**Target / Gene**
HIVR1R	CAGGATTRACTGCGAATCG	*gag*
HIVR2R	YACACAATAGAGGGTTGCT	*gag*
HIVR3R	TTCTAGGTGATATGGCYTGA	*gag*
HIVR4R	CATTTGCATRGCTGCTT	*gag*
HIVR5R	TTTCTCCTACTGGGATAGGT	*gag*
HIVR6R	TTACCTCYTGTGAAGCT	*gag*
HIVR7R	GGTTCCTAAAATTGCCTYTC	*gag*
HIVR8R	CCTTCCYTTCCACATYTCC	*gag*
HIVR9R	GTGACGAGGGGTCGTT	*gag-pol*
HIVR10R	GGTCCTACTARTACTGTACC	*pol*
HIVR11R	TAACTYTTGGGCCATCC	*pol*
HIVR12R	CCTGCRGGATGTGGTA	*pol*
HIVR13R	GGAATATTGCTGGTGATCC	*pol*
HIVR14R	TTGTCTGGTGTGGTAAAYC	*pol*
HIVR15R	CTGSRTAAATCTGACTTGC	*pol*
HIVR16R	GGGTCATAATAYACTCCAYG	*pol*
HIVR17R	CCAATACTCYRTCCACCAT	*pol*
HIVR18R	CCGAATCCTGCAARGCT	*pol*
HIVR19R	TKATCTGGTTGTGCYTG	*pol*

HIVR20R	CTGATTCCAGYACTGACT	*pol*
HIVR21R	CATGCATGGCTTCTCCTT	*pol*
HIVR22R	TTGCTGCCATTGTCWGT	*pol*
HIVR23R	AGATGTTCAGCCTGATC	*pol*
HIVR24R	GCTGTCYCTGTAATAAACC	*pol*
HIVR25R	CTGTCTACTTGCCACACA	*pol*
HIVR26R	CTGTATGCAGACCCCAA	*pol*
HIVR27R	AGGGTCTACTTGTGTGCT	*vif*
HIVR28R	TCCCTCTGYGGCCCTT	*vif*
HIVR29R	GCTCTAGTCTAGGATCTACTG	*vpr*
HIVR30R	GAGAARCTTGATGAGTCTGA	*vpr-vpu*
HIVR31R	CTYTCATTGCCACTGTCT	*vpu-env*
HIVR32R	GTGGGTTGGGGTCTGT	*env*
HIVR33R	CCTTAYCTCTTATKCTTGTG	*env*
HIVR34R	ATGGGAATTGGCTCAAAKG	*env*
HIVR35R	GTTGTTGGGTCTTRTAC	*env*
HIVR36R	ACAAATGCTCTCCCTGGT	*env*
HIVR37R	CCCCTCCACAATTAAAACTG	*env*
HIVR38R	CTTCTCCAATTGTCCCTCA	*env*
HIVR39R	TGGYCTGTACCGTCAG	*env*
HIVR40R	TGAGTTTTCCAGAGCAACC	*env*
HIVR41R	TAAACCTAYCAAGCCTCCT	*env*
HIVR42R	GGATCTGTCTCTGTCTCTC	*env*
HIVR43R	GCGTCCCAGAAGTTCCA	*env*
HIVR44R	CTTTCCAAGCCCTGTCT	*env*
HIVR45R	TAGCCAGGCACAAKCA	*nef*
HIVR46R	GTRAAYTAGCCCTTCCAGT	*nef*-3’ LTR

Table A.4. Primers designed using PrimerDesign-M to amplify areas with low-coverage within *pol*.

Forward Primer	Sequence (5' → 3')
A3_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTATGATCARATASYYATAGAAATYTG
A3_Set5_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTATGATCARATASYYATAGAAATYTG
B3_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCTGTACCAGTAAAATTRAAGCCAG
B3_Set5_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGACTGTACCAGTAAAATTRAAGCC
B4_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCTGGACTGTCAATGAYATACAG
B4_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCTGGACTGTCAATGAYATAC
C2_Set2_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGAGAAGARATGATGACAGCATG
C2_Set3_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGGAAGARATGATGACAGCATG
C4_Set2_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTATGACCCATCAAAAGAYTTRATAG
C4_Set5_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGGACCCATCAAAAGAYTTRATAGC
C5_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGACTGYAGTCCAGGRATATGGC
C5_Set5_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCTGYAGTCCAGGRATATGGC
D5_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTAGCAGGAAGATGGCCAG
D5_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTAGCAGGAAGATGGCCAG
E2-1_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

	NNNNGGTCACTCTTTGGCARCGACC
E2-1_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTCACTCTTTGGCARCGACC
E2-2_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGACAGGAGCAGATGATACAG
E2-2_Set3_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGACAGGAGCAGATGATACAG
E4_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGARTTTGTYAAYACCCCTC
E4_Set4_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGGGARTTTGTYAAYACCCCTCC
E5_Set1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGGCTGARCAYCTYAARACAG
E5_Set2_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCTGARCAYCTYAARACAGC
A3_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTTTTCAGGCCCAATTYTTG
A3_Set5_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTTTCAGGCCCAATTYTTG
B3_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGGTATYCCTARYTGAACTTCCC
B3_Set5_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGGTATYCCTARYTGAACTTCCC
B4_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGAATGGYTCYTGRTAAATTTG
B4_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGAATGGYTCYTGRTAAATTTG
C2_Set2_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGGAAGGCCARATYYTCCC
C2_Set3_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGGAAGGCCARATYYTCCC
C4_Set2_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGACAAAYTCCCAYTCAGGAATYC

C4_Set5_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGACAAAYTCCCAYTCAGGAATYC
C5_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGACTTTGGGGATTGTAGGG
C5_Set5_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTGACTTTGGGGATTGTAGGG
D5_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTCTAYTATTCTTTCCCCTGC
D5_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTGCTATTATRTCTAYTATTCTTTCC
E2-1_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTATGTTGACAGGTGTAGGTC
E2-1_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGATGTTGACAGGTGTAGGTCC
E2-2_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGCTGGCTTYAATTTTACTGGTACAG
E2-2_Set3_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTGGCTTYAATTTTACTGGTACAG
E4_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTRACTAWYTCTGATTCAC
E4_Set4_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGATTTGRYTRACTAWYTCTGATTCAC
E5_Set1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTAATCYTCATCCTGTCTAC
E5_Set2_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTAATCYTCATCCTGTCTAC
D5.1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTAGCAGGRRGATGGCCAG
D5.2_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGTAGCAGGRRGATGGCCAG
B5.1_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGATTGGRGGAAATGAAMARGTRG
B5.3_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT

	NNNNGGATTGGRGGAAATGAAMARGTRG
D5.1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTCTGCTGTCYCTGTAAYARACC
D5.2_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTGCTGTCYCTGTAAYARACCCG
B5.1_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTKACTGGCCATCYYCCTGC
B5.3_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGTKACTGGCCATCYYCCTG
E3.2_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCCACARGGRTGGAAAGG
E3.3_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGCCACARGGRTGGAAAGG
A5.3_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGAYTCMCARTATGCATTAGG
A5.5_Forward_NGS	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCGATCT NNNNGGACAGAYTCMCARTATGCATTAGG
E3.2_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGACTGTCCAYTTRTCAGGATG
E3.3_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGCTGTCCAYTTRTCAGGATG
A5.3_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGGCTACTATTTCYTTKGS
A5.5_Reverse_NGS	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT NNNNNNNNNNGGCARCWRGCTACTATTTCYTTKG

Table A.5. List of primers used in the finalized *pol* gene direct-sequencing primer protocol.

Primer	5’ – 3’ Sequence	Comment
HIVM_A3.4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TTTTCAGGCCCAATTYTTG	cDNA primer with UMI for PR
HIVM_A4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG CYAGTTCTAGCTCTGCTTC	cDNA primer with UMI for RT
HIVM_A5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TCTTGGGCCTTATCTATYCC	cDNA primer with UMI for RT
HIVM_A6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG GTGGGAYRTGTACTTCTGA	cDNA primer with UMI for IN
HIVM_B3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TGAACTTCCCAGAARYC	cDNA primer with UMI for PR
HIVM_B4.1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG AATGGYTCYTGRTAAATTTG	cDNA primer with UMI for RT
HIVM_B5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG GCTACCAGRATAAYTTTTCC	cDNA primer with UMI for RT
HIVM_B6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG RCCCAAATGCCAGTCYCT	cDNA primer with UMI for IN
HIVM_C2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG AAGGCCAGATYTTCCCTA	cDNA primer with UMI for PR

HIVM_C3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG GCAGTATACTTYCTRAAGTC	cDNA primer with UMI for RT
HIVM_C4.2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG ACAAAYTCCCAYTCAGGAATYC	cDNA primer with UMI for RT
HIVM_C5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG GATYCCYGCCCACCAA	cDNA primer with UMI for IN
HIVM_D2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG CYTTTAGTTGCCCYCCTA	cDNA primer with UMI for PR
HIVM_D3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TGYYGCCCTATTTCTAAGTC	cDNA primer with UMI for RT
HIVM_D4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG CTRTTRGCTGCCCCATC	cDNA primer with UMI for RT
HIVM_D5.4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TGCTATTATRTCTAYTATTCTTTCC	cDNA primer with UMI for IN
HIVM_E2-1.1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TATGTTGACAGGTGTAGGTC	cDNA primer with UMI for PR
HIVM_E3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG GGATGGAGTTCATAACCCAT	cDNA primer with UMI for RT
HIVM_E4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG TKATCTGGYTGTGCTTG	cDNA primer with UMI for RT
HIVM_E5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GG ACCTGCCRTCTGTTTTCC	cDNA primer with UMI for IN

HIVM_A3F	ATAGGGGGAATTGGAGGKT	1st round PCR forward primer for PR
HIVM_A4F	TCAGAARGAACCYCCATTCC	1st round PCR forward primer for RT
HIVM_A5F	CAGGATTCRGGAYTAGAAG	1st round PCR forward primer for RT
HIVM_A6F	AGGGGCAGTAGTAATACAAG	1st round PCR forward primer for IN
HIVM_B2F	GGATGACAGAAACCTTGTTG	1st round PCR forward primer for PR
HIVM_B3F	AAAGCCAGGAATGGATG	1st round PCR forward primer for RT
HIVM_B4F	CCTTAGRGGARCCAAAGCA	1st round PCR forward primer for RT
HIVM_B5F	RCTGGAATCAGGAAAGTAC	1st round PCR forward primer for IN
HIVM_C2F	TGTCAGGGAGTRGGRGGA	1st round PCR forward primer for PR
HIVM_C3F	TGGGCCTGAAAAYCCAT	1st round PCR forward primer for RT
HIVM_C4F	GGAGTGTATTATGACCCATC	1st round PCR forward primer for RT

HIVM_C5F	YCCAGGAATATGGCAAYTAG	1st round PCR forward primer for IN
HIVM_D2F	WTGYACTGAGAGACAGGCTA	1st round PCR forward primer for PR
HIVM_D3F	GTAYTGGATGTRGGTGATG	1st round PCR forward primer for RT
HIVM_D4F	TTAACAGAGGCAGTGCA	1st round PCR forward primer for RT
HIVM_D5F	GCAGGAAGATGGCCAGT	1st round PCR forward primer for IN
HIVM_E2F	CAGRTCACTCTTTGGCAA	1st round PCR forward primer for PR
HIVM_E3F	AGGATCACCAGCAATATTCC	1st round PCR forward primer for RT
HIVM_E4F	GARAAAGAACCCATARTAGG	1st round PCR forward primer for RT
HIVM_E5F	AAAGAAAAGGGGGGATTGG	1st round PCR forward primer for IN
Universal Adapter Reverse	GTGACTGGAGTTCAGACGTGTGCTC	1st round PCR reverse primer
HIVM_A3.4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTATGATCARATASYYATAGAAATYTG	2nd round PCR forward primer for PR w/ Illumina adapter sequence

HIVM_A4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTCAGAARGAACCYCCATTCC	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_A5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGCAGGATTCRGGAYTAGAAG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_A6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGAGGGGCAGTAGTAATACAAG	2nd round PCR forward primer for IN w/ Illumina adapter sequence
HIVM_B3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGAAAGCCAGGAATGGATG	2nd round PCR forward primer for PR w/ Illumina adapter sequence
HIVM_B4.1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGCTGGACTGTCAATGAYATACAG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_B5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGRCTGGAATCAGGAAAGTAC	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_B6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGGCAGGTGATGATWGTGTG	2nd round PCR forward primer for IN w/ Illumina adapter sequence
HIVM_C2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTGTCAGGGAGTRGGRGGA	2nd round PCR forward primer for PR w/ Illumina adapter sequence
HIVM_C3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTGGGCCTGAAAAYCCAT	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_C4.2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTATGACCCATCAAAAGAYTTRATAG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_C5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGYCCAGGAATATGGCAAYTAG	2nd round PCR forward primer for IN w/ Illumina adapter sequence

HIVM_D2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGWTGYACTGAGAGACAGGCTA	2nd round PCR forward primer for PR w/ Illumina adapter sequence
HIVM_D3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGGTAYTGGATGTRGGTGATG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_D4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTTAACAGAGGCAGTGCA	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_D5.4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTAGCAGGAAGATGGCCAG	2nd round PCR forward primer for IN w/ Illumina adapter sequence
HIVM_E2-1.1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGTCACTCTTTGGCARCGACC	2nd round PCR forward primer for PR w/ Illumina adapter sequence
HIVM_E3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGAGGATCACCAGCAATATTCC	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_E4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGGARAAAGAACCCATARTAGG	2nd round PCR forward primer for RT w/ Illumina adapter sequence
HIVM_E5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACG CTCTTCCGATCTNNNNGGAAAGAAAAGGGGGGATTGG	2nd round PCR forward primer for IN w/ Illumina adapter sequence
Indexed Adapter[c]	CAAGCAGAAGACGGCATACGAGAT**NNNNNN**GTGACTGGAGTTCAG ACGTGTGCTC	2nd round PCR reverse primer w/ Illumina adapter sequence and barcodes

Table A.6. Complete list of HIVM panel Option #1 designed direct sequencing primers for reverse transcription.

Primer Name	Sequence (5' → 3')	Target gene
HIVM_A1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGYTCCCTGCTTGCCCAT	*gag*
HIVM_A2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTCTGGGTTYGCATTTTGG	*gag*
HIVM_A3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTTCTGTYAATGGCCAT	*pol*
HIVM_A4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCYAGTTCTAGCTCTGCTTC	*pol*
HIVM_A5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTCTTGGGCCTTATCTATYCC	*pol*
HIVM_A6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGTGGGAYRTGTACTTCTGA	*pol*
HIVM_A7R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGGCTCTARTYTAGGATCTAC	*vpr*
HIVM_A8R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTGAGGTRTTACAAYTTATC	*env*
HIVM_A9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTRATGGGAGGRGCAT	*env*
HIVM_A10R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTCTCTCTCCACCTTCTYC	*env*
HIVM_B1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTAARGCTTCCTTGGTGTC	*gag*
HIVM_B2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTGYCCTTCTTTGCCACA	*gag*
HIVM_B3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTGAACTTCCCAGAARYC	*pol*
HIVM_B4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCACCCCTCRTYCTTGCAT	*pol*
HIVM_B5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGCTACCAGRATAAYTTTTCC	*pol*
HIVM_B6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGRCCCAAATGCCAGTCYCT	*pol*
HIVM_B7R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGMTCTKCGTCGCTGTCTC	*vpr*
HIVM_B8R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGTATGGGAATTGGCTSA	*env*
HIVM_B9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTYTGCACCACTCTTCTCT	*env*
HIVM_B10R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTRTCTGTCCCCTCAGYT	*env*
HIVM_C1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCTGAAAGCCTTYTCTTCTAC	*gag*
HIVM_C2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGAAGGCCAGATYTTCCCTA	*pol*
HIVM_C3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGCAGTATACTTYCTRAAGTC	*pol*
HIVM_C4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTCCMCCATGYTTCCCATG	*pol*
HIVM_C5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGGATYCCYGCCCACCAA	*pol*
HIVM_C6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGCCTAGGACTAACTMTAYGTC	*vif*
HIVM_C7R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNNN**GGTGCCACTGTCTTCTGCT	*vpu*

HIVM_C8R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCAGATTYRTTCAGCTGTAC	*env*
HIVM_C9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTGTTGCGCYTCAATAGC	*env*
HIVM_C10R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGACCAYTTGCCACCCAT	*env*
HIVM_D1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGAAYAGGCCCTGCATGCA	*gag*
HIVM_D2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCYTTTAGTTGCCCYCCTA	*pol*
HIVM_D3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTGYYGCCCTATTTCTAAGTC	*pol*
HIVM_D4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCTRTTRGCTGCCCCATC	*pol*
HIVM_D5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTTCCCCTGCACTGTAYC	*pol*
HIVM_D6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCTTGTTCCATCTATCCTCTG	*vif*
HIVM_D7R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCTTCYTTCCACACAGGTAC	*env*
HIVM_D8R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGATGCTCTCCCTGGTCCT	*env*
HIVM_D9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCCAAGGCACAGYAGTG	*env*
HIVM_D10R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTGGTCTTAAAGGWACCTGRG	*nef*
HIVM_E1R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGGTTCYTTTGGTCCTTGTYT	*gag*
HIVM_E2R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTTGACAGGTGTAGGTCCT	*pol*
HIVM_E3R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGGGATGGAGTTCATAACCCAT	*pol*
HIVM_E4R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTKATCTGGYTGTGCTTG	*pol*
HIVM_E5R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGACCTGCCRTCTGTTTTCC	*pol*
HIVM_E6R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTTATGGCYTCCACTCCT	*vpr*
HIVM_E7R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGCACATGGCTTTARGCTT	*env*
HIVM_E8R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGYTTCTGGGTCCCCYCCT	*env*
HIVM_E9R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGTCCACAAACTTGCCCAT	*env*
HIVM_E10R	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCT**NNNNNNNNN**GGGWAGCACCAYCCAAAGGT	*nef*

Table A.7. Complete list of HIVM panel Option #1 designed direct sequencing primers for final amplification (2nd round PCR).

Primer Name	Sequence (5' → 3')	Target gene
HIVM_A1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCCTCAGACCMTTTTAGTCA	*gag*
HIVM_A2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCACCTATCCCAGTAGGAGA	*gag*
HIVM_A3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGATAGGGGGAATTGGAGGKT	*pol*
HIVM_A4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTCAGAARGAACCYCCATTCC	*pol*
HIVM_A5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCAGGATTCRGGAYTAGAAG	*pol*
HIVM_A6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAGGGGCAGTAGTAATACAAG	*pol*
HIVM_A7F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGAGGAGCTTAAGARTGAAGC	*vpr*
HIVM_A8F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGRAYATGGTAGAACAGATGC	*env*
HIVM_A9F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGGACCCAGAARTTGTAAYG	*env*
HIVM_A10F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGATCGCARAACCARCAAG	*env*
HIVM_B1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGYTAAGGCCAGGRGGAA	*gag*
HIVM_B2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGGATGACAGAAACCTTGTTG	*gag*
HIVM_B3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAAAGCCAGGAATGGATG	*pol*
HIVM_B4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCCTTAGRGGARCCAAAGCA	*pol*
HIVM_B5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGRCTGGAATCAGGAAAGTAC	*pol*

HIVM_B6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGCAGGTGATGATWGTGTG	*pol*
HIVM_B7F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTGYCRACATAGCAGAATAGG	*vpr*
HIVM_B8F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGYTAACCCCACTCTGTGT	*env*
HIVM_B9F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGACATGTGGCAGAAAGTAGG	*env*
HIVM_B10F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGACAGGCCCGAAGGAA	*env*
HIVM_C1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGRCCCTCTATTGTGTRCATC	*gag*
HIVM_C2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTGTCAGGGAGTRGGRGGA	*pol*
HIVM_C3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTGGGCCTGAAAAYCCAT	*pol*
HIVM_C4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGGAGTGTATTATGACCCATC	*pol*
HIVM_C5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGYCCAGGAATATGGCAAYTAG	*pol*
HIVM_C6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTAGRGGATGCTARATTGGT	*vif*
HIVM_C7F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGSCTTAGGCATCTCCYATGG	*vpu*
HIVM_C8F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGMTTACACARGCCTGTCCA	*env*
HIVM_C9F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGGGACAATTGGAGAAGTGA	*env*
HIVM_C10F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGAGGAYTGTGGAAMTTCTG	*env*
HIVM_D1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGARGCMATATCACCTAGAAC	*gag*

HIVM_D2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGWTGYACTGAGAGACAGGCTA	*pol*
HIVM_D3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGTAYTGGATGTRGGTGATG	*pol*
HIVM_D4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTTAACAGAGGCAGTGCA	*pol*
HIVM_D5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGCAGGAAGATGGCCAGT	*pol*
HIVM_D6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAGTCKCCRTAGAATGGAGGA	*vif*
HIVM_D7F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGCAATAGTTGTGTGGWCYA	*env*
HIVM_D8F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCASCACAGTACAATGTACAC	*env*
HIVM_D9F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGACGGTACAGRCCAGA	*env*
HIVM_D10F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGACAGGGCYTRGAAAGG	*nef*
HIVM_E1F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTCAATGARGAAGCTGCA	*gag*
HIVM_E2F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGCAGRTCACTCTTTGGCAA	*pol*
HIVM_E3F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAGGATCACCAGCAATATTCC	*pol*
HIVM_E4F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGARAAAGAACCCATARTAGG	*pol*
HIVM_E5F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAAAGAAAAGGGGGGATTGG	*pol*
HIVM_E6F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAAAGCCRCCTTTGCCTA	*vpr*
HIVM_E7F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAAWTGTGGGTCACAGTC	*env*

HIVM_E8F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGGTAYAAGACCCAACAAC	*env*
HIVM_E9F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGAGARTCCTGGCTGTRGAA	*env*
HIVM_E10F	AATGATACGGCGACCACCGAGATCTACACTCTTTCCCTACACGACGCTCTTCCG ATCTNNNNGGTGGCTAGAAGCACAAGAG	*nef*

Table A.8. List of single index barcode sequences used in the direct sequencing protocol final amplification (2nd round PCR).

Primer Name	Sequence (5' → 3')[a]
Index 1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCAGACGTGTGCTC
Index 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCAGACGTGTGCTC
Index 3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCAGACGTGTGCTC
Index 4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCAGACGTGTGCTC
Index 5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCAGACGTGTGCTC
Index 6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCAGACGTGTGCTC
Index 7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCAGACGTGTGCTC
Index 8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTC
Index 9	CAAGCAGAAGACGGCATACGAGATCTGATCGTGACTGGAGTTCAGACGTGTGCTC
Index 10	CAAGCAGAAGACGGCATACGAGATAAGCTAGTGACTGGAGTTCAGACGTGTGCTC
Index 11	CAAGCAGAAGACGGCATACGAGATGTAGCCGTGACTGGAGTTCAGACGTGTGCTC
Index 12	CAAGCAGAAGACGGCATACGAGATTACAAGGTGACTGGAGTTCAGACGTGTGCTC
Index 13	CAAGCAGAAGACGGCATACGAGATTTGACTGTGACTGGAGTTCAGACGTGTGCTC
Index 14	CAAGCAGAAGACGGCATACGAGATGGAACTGTGACTGGAGTTCAGACGTGTGCTC
Index 15	CAAGCAGAAGACGGCATACGAGATTGACATGTGACTGGAGTTCAGACGTGTGCTC
Index 16	CAAGCAGAAGACGGCATACGAGATGGACGGGTGACTGGAGTTCAGACGTGTGCTC
Index 18	CAAGCAGAAGACGGCATACGAGATGCGGACGTGACTGGAGTTCAGACGTGTGCTC
Index 19	CAAGCAGAAGACGGCATACGAGATTTTCACGTGACTGGAGTTCAGACGTGTGCTC
Index 20	CAAGCAGAAGACGGCATACGAGATGGCCACGTGACTGGAGTTCAGACGTGTGCTC
Index 21	CAAGCAGAAGACGGCATACGAGATCGAAACGTGACTGGAGTTCAGACGTGTGCTC
Index 22	CAAGCAGAAGACGGCATACGAGATCGTACGGTGACTGGAGTTCAGACGTGTGCTC
Index 23	CAAGCAGAAGACGGCATACGAGATCCACTCGTGACTGGAGTTCAGACGTGTGCTC
Index 25	CAAGCAGAAGACGGCATACGAGATATCAGTGTGACTGGAGTTCAGACGTGTGCTC
Index 27	CAAGCAGAAGACGGCATACGAGATAGGAATGTGACTGGAGTTCAGACGTGTGCTC
Index 150	CAAGCAGAAGACGGCATACGAGATTAGTTCGTGACTGGAGTTCAGACGTGTGCTC
Index 151	CAAGCAGAAGACGGCATACGAGATTTAATGGTGACTGGAGTTCAGACGTGTGCTC
Index 152	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCAGACGTGTGCTC
Index 153	CAAGCAGAAGACGGCATACGAGATCACACAGTGACTGGAGTTCAGACGTGTGCTC

Index 154	CAAGCAGAAGACGGCATACGAGATGTCAGGGTGACTGGAGTTCAGACGTGTGCTC
Index 155	CAAGCAGAAGACGGCATACGAGATTTGAACGTGACTGGAGTTCAGACGTGTGCTC
Index 156	CAAGCAGAAGACGGCATACGAGATATATGTGTGACTGGAGTTCAGACGTGTGCTC
Index 157	CAAGCAGAAGACGGCATACGAGATATGTGCGTGACTGGAGTTCAGACGTGTGCTC
Index 158	CAAGCAGAAGACGGCATACGAGATCCGATAGTGACTGGAGTTCAGACGTGTGCTC
Index 159	CAAGCAGAAGACGGCATACGAGATCAGGAAGTGACTGGAGTTCAGACGTGTGCTC
Index 160	CAAGCAGAAGACGGCATACGAGATGCATAAGTGACTGGAGTTCAGACGTGTGCTC
Index 161	CAAGCAGAAGACGGCATACGAGATATCGTAGTGACTGGAGTTCAGACGTGTGCTC
Index 162	CAAGCAGAAGACGGCATACGAGATACTAAAGTGACTGGAGTTCAGACGTGTGCTC
Index 163	CAAGCAGAAGACGGCATACGAGATTAGGAAGTGACTGGAGTTCAGACGTGTGCTC
Index 164	CAAGCAGAAGACGGCATACGAGATGTCCGTGTGACTGGAGTTCAGACGTGTGCTC
Index 165	CAAGCAGAAGACGGCATACGAGATCGAGGCGTGACTGGAGTTCAGACGTGTGCTC
Index 166	CAAGCAGAAGACGGCATACGAGATTTCCAAGTGACTGGAGTTCAGACGTGTGCTC
Index 167	CAAGCAGAAGACGGCATACGAGATAACTCAGTGACTGGAGTTCAGACGTGTGCTC
Index 168	CAAGCAGAAGACGGCATACGAGATTCAGACGTGACTGGAGTTCAGACGTGTGCTC
Index 169	CAAGCAGAAGACGGCATACGAGATCCACGAGTGACTGGAGTTCAGACGTGTGCTC
Index 170	CAAGCAGAAGACGGCATACGAGATGCGTGCGTGACTGGAGTTCAGACGTGTGCTC
Index 171	CAAGCAGAAGACGGCATACGAGATCAAAGGGTGACTGGAGTTCAGACGTGTGCTC
Index 172	CAAGCAGAAGACGGCATACGAGATTGCGGCGTGACTGGAGTTCAGACGTGTGCTC
Index 173	CAAGCAGAAGACGGCATACGAGATCGCTGTGTGACTGGAGTTCAGACGTGTGCTC

[a]Barcode sequence is a string of six-nucleotides that differ from one another by 1 or more bases.

References

1. World Health Organization. HIV and AIDS. [cited 1 Nov 2023]. Available: https://www.who.int/news-room/fact-sheets/detail/hiv-aids
2. Santos AF, Soares MA. HIV Genetic Diversity and Drug Resistance. Viruses. 2010;2: 503. doi:10.3390/V2020503
3. World Health Organization. HIV. 2022 [cited 14 Dec 2022]. Available: https://www.who.int/data/gho/data/themes/hiv-aids
4. Public Health Agency of Canada. HIV in Canada - People living with HIV and new HIV infections, 2020 - Canada.ca. [cited 16 Dec 2022]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-canada-people-living-with-hiv-new-infections-2020.html
5. Public Health Agency of Canada. HIV in Canada: 2021 surveillance highlights - Canada.ca. [cited 16 Dec 2022]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-2021-surveillance-highlights.html
6. Public Health Agency of Canada. HIV in Canada: 2022 Surveillance highlights - Canada.ca. [cited 4 Dec 2023]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-2022-surveillance-highlights.html
7. German Advisory Committee Blood (Arbeitskreis Blut) S ‘Assessment of PT by B. Human Immunodeficiency Virus (HIV). Transfus Med Hemotherapy. 2016;43: 203–222. doi:10.1159/000445852
8. Nastri BM, Pagliano P, Zannella C, Folliero V, Masullo A, Rinaldi L, et al. HIV and Drug-Resistant Subtypes. Microorganisms. 2023;11: 221. doi:10.3390/MICROORGANISMS11010221/S1

9. Worobey M, Gemmel M, Teuwen DE, Haselkorn T, Kunstman K, Bunce M, et al. Direct Evidence of Extensive Diversity of HIV-1 in Kinshasa by 1960. Nature. 2008;455: 661. doi:10.1038/NATURE07390

10. Sharp PM, Hahn BH. Origins of HIV and the AIDS Pandemic. [cited 1 Nov 2023]. doi:10.1101/cshperspect.a006841

11. Merson MH, O'Malley J, Serwadda D, Apisuk C. The history and challenge of HIV prevention. Lancet. 2008;372: 475–488. doi:10.1016/S0140-6736(08)60884-3

12. Williams A, Menon S, Crowe M, Agarwal N, Biccler J, Bbosa N, et al. Geographic and Population Distributions of Human Immunodeficiency Virus (HIV)–1 and HIV-2 Circulating Subtypes: A Systematic Literature Review and Meta-analysis (2010–2021). J Infect Dis. 2023;228: 1583–1591. doi:10.1093/INFDIS/JIAD327

13. Bbosa N, Kaleebu P, Ssemwanga D. HIV subtype diversity worldwide. Curr Opin HIV AIDS. 2019;14: 153–160. doi:10.1097/COH.0000000000000534

14. International Agency for Research on Cancer. Human Immunodeficiency Virus-1. 100B ed. Biological Agents. 100B ed. Lyon (FR): International Agency for Research on Cancer; 2012. pp. 215–218. Available: https://www.ncbi.nlm.nih.gov/books/NBK304351/#a010.R113

15. Engelman A, Cherepanov P. The structural biology of HIV-1: mechanistic and therapeutic insights. Nat Rev Microbiol 2012 104. 2012;10: 279–290. doi:10.1038/nrmicro2747

16. Hemelaar J, Loganathan S, Elangovan R, Yun J, Dickson-Tetteh L, Kirtley S. Country Level Diversity of the HIV-1 Pandemic between 1990 and 2015. J Virol. 2020;95. doi:10.1128/JVI.01580-20

17. Clavel F, Hance AJ. HIV Drug Resistance. N Engl J Med. 2004;350: 1023–1035.

doi:10.1056/NEJMRA025195

18. Mansky LM. HIV mutagenesis and the evolution of antiretroviral drug resistance. Drug Resist Updat. 2002;5: 219–223. doi:10.1016/S1368-7646(02)00118-8

19. Rhee S-Y, Gonzales MJ, Kantor R, Betts BJ, Ravela J, Shafer RW. Human immunodeficiency virus reverse transcriptase and protease sequence database. Nucleic Acids Res. 2003;31: 298–303. doi:10.1093/nar/gkg100

20. Shafer RW. Rationale and Uses of a Public HIV Drug-Resistance Database. J Infect Dis. 2006;194: 51–58. Available: http://hivdb.stanford.edu/pages/genotype-rx.html

21. Wensing AM, Calvez V, Ceccherini-Silberstein F, Charpentier C, Günthard HF, Paredes R, et al. 2022 Update of the Drug Resistance Mutations in HIV-1. Top Antivir Med. 2022;30: 559. Available: /pmc/articles/PMC9681141/

22. Hunter JR, Matos Dos Santos DE, Munerato P, Janini LM, Castelo A, Sucupira MC, et al. Antiretroviral Drug-Resistance Mutations on the Gag Gene: Mutation Dynamics during Analytic Treatment Interruption among Individuals Experiencing Virologic Failure. Pathogens. 2022;11: 534. doi:10.3390/pathogens11050534

23. Van Duyne R, Kuo LS, Pham P, Fujii K, Freed EO. Mutations in the HIV-1 envelope glycoprotein can broadly rescue blocks at multiple steps in the virus replication cycle. Proc Natl Acad Sci U S A. 2019;116: 9040–9049. doi:10.1073/PNAS.1820333116/SUPPL_FILE/PNAS.1820333116.SAPP.PDF

24. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. Clin Infect Dis. 2006;42: 1608–1618. doi:10.1086/503914

25. Nykoluk M, Taylor T, Ji H, Enns E. HyDRA web user guide. In: Public Health Agency of Canada, Ottawa, Canada. [Internet]. 2016. Available:

https://hydra.canada.ca/HyDRA_Web_User_Guide_Final_6Sept2016.pdf

26. Taylor T, Lee ER, Nykoluk M, Enns E, Liang B, Capina R, et al. A MiSeq-HyDRA platform for enhanced HIV drug resistance genotyping and surveillance. Sci Rep. 2019;9. doi:10.1038/S41598-019-45328-3

27. Ji H, Sandstrom P. Overview of the Analytes Applied in Genotypic HIV Drug Resistance Testing. Pathogens. 2022;11. doi:10.3390/PATHOGENS11070739

28. World Health Organization. WHO/HIVRESNET HIV DRUG RESISTANCE LABORATORY OPERATIONAL FRAMEWORK HIV DRUG RESISTANCE. 2017.

29. Parry CM, Parkin N, Diallo K, Mwebaza S, Batamwita R, DeVos J, et al. Field study of dried blood spot specimens for HIV-1 drug resistance genotyping. J Clin Microbiol. 2014;52: 2868–2875. doi:10.1128/JCM.00544-14

30. Rodriguez RJ, Dayhoff DE, Chang G, Cassol SA, Birx DL, Artenstein AW, et al. Comparison of Serum and Plasma Viral RNA Measurements in Primary and Chronic Human Immunodeficiency Virus Type 1 Infection. J Acquir Immune Defic Syndr Hum Retrovirology. 1997;15: 49–53. Available: https://journals.lww.com/jaids/fulltext/1997/05010/comparison_of_serum_and_plasma_viral_rna.8.aspx

31. García-Lerma JG, McNulty A, Jennings C, Huang D, Heneine W, Bremer JW. Rapid decline in the efficiency of HIV drug resistance genotyping from dried blood spots (DBS) and dried plasma spots (DPS) stored at 37°C and high humidity. J Antimicrob Chemother. 2009;64: 33–36. doi:10.1093/JAC/DKP150

32. Rottinghaus EK, Beard RS, Bile E, Modukanele M, Maruping M, Mine M, et al. Evaluation of Dried Blood Spots Collected on Filter Papers from Three Manufacturers

Stored at Ambient Temperature for Application in HIV-1 Drug Resistance Monitoring. PLoS One. 2014;9: e109060. doi:10.1371/JOURNAL.PONE.0109060

33. Aitken SC, Wallis CL, Stevens W, De Wit TR, Schuurman R. Stability of HIV-1 Nucleic Acids in Dried Blood Spot Samples for HIV-1 Drug Resistance Genotyping. PLoS One. 2015;10. doi:10.1371/JOURNAL.PONE.0131541

34. World Health Organization. WHO manual for HIV drug resistance testing using dried blood spot specimens. Third. Geneva, Switzerland; 2020. Available: https://www.who.int/publications/i/item/9789240009424

35. Lim MD. Dried Blood Spots for Global Health Diagnostics and Surveillance: Opportunities and Challenges. Am J Trop Med Hyg. 2018;99: 256. doi:10.4269/AJTMH.17-0889

36. Tuaillon E, Kania D, Pisoni A, Bollore K, Taieb F, Ontsira Ngoyi EN, et al. Dried Blood Spot Tests for the Diagnosis and Therapeutic Monitoring of HIV and Viral Hepatitis B and C. Front Microbiol. 2020;11. doi:10.3389/FMICB.2020.00373

37. Stranneheim H, Lundeberg J. Stepping stones in DNA sequencing. Biotechnol J. 2012;7: 1063–1073. doi:10.1002/BIOT.201200153

38. Manyana S, Gounder L, Pillay M, Manasa J, Naidoo K, Chimukangara B. HIV-1 Drug Resistance Genotyping in Resource Limited Settings: Current and Future Perspectives in Sequencing Technologies. Viruses 2021, Vol 13, Page 1125. 2021;13: 1125. doi:10.3390/V13061125

39. Tekin D, Gokengin D, Onay H, Erensoy S, Sertoz R. Investigation of drug resistance against protease, reverse transcriptase, and integrase inhibitors by next-generation sequencing in HIV-positive patients. J Med Virol. 2021;93: 3627–3633.

doi:10.1002/JMV.26582

40. Becker MG, Liang D, Cooper B, Le Y, Taylor T, Lee ER, et al. Development and application of performance assessment criteria for next-generation sequencing-based HIV drug resistance assays. Viruses. 2020;12. doi:10.3390/v12060627

41. Gryseels S, Watts TD, Mpolesha JMK, Larsen BB, Lemey P, Muyembe-Tamfum JJ, et al. A near full-length HIV-1 genome from 1966 recovered from formalin-fixed paraffin-embedded tissue. Proc Natl Acad Sci U S A. 2020;117. doi:10.1073/PNAS.1913682117/-/DCSUPPLEMENTAL

42. Worobey M, Watts TD, McKay RA, Suchard MA, Granade T, Teuwen DE, et al. 1970s and 'Patient 0' HIV-1 genomes illuminate early HIV/AIDS history in North America. Nature. 2016;539: 98–101. doi:10.1038/NATURE19827

43. Fitzpatrick AH, Rupnik A, O'Shea H, Crispie F, Keaveney S, Cotter P. High Throughput Sequencing for the Detection and Characterization of RNA Viruses. Front Microbiol. 2021;12: 190. doi:10.3389/fmicb.2021.621719

44. Bonsall D, Golubchik T, de Cesare M, Limbada M, Kosloff B, MacIntyre-Cockett G, et al. A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. J Clin Microbiol. 2020;58: 382–402. doi:10.1128/JCM.00382-20

45. Zakotnik S, Knap N, Bogovič P, Zorec TM, Poljak M, Strle F, et al. Complete Genome Sequencing of Tick-Borne Encephalitis Virus Directly from Clinical Samples: Comparison of Shotgun Metagenomic and Targeted Amplicon-Based Sequencing. Viruses. 2022;14: 1267. doi:10.3390/V14061267/S1

46. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes

directly from clinical samples. Nat Protoc. 2017;12: 1261–1277. doi:10.1038/NPROT.2017.066

47. Arana C, Liang C, Brock M, Zhang B, Zhou J, Chen L, et al. A short plus long-amplicon based sequencing approach improves genomic coverage and variant detection in the SARS-CoV-2 genome. PLoS One. 2022;17. doi:10.1371/JOURNAL.PONE.0261014

48. Park C, Kim KW, Park D, Hassan Z ul, Park EC, Lee CS, et al. Rapid and sensitive amplicon-based genome sequencing of SARS-CoV-2. Front Microbiol. 2022;13: 876085. doi:10.3389/FMICB.2022.876085/BIBTEX

49. Woolworth Hirschhorn J, Schandl CA, Nolte FS. Chapter 69 – Polymerase Chain Reaction And Other Nucleic Acid Amplification Technology. Henry's Clin Diagnosis Manag by Lab Methods. 2022; 1387–1400.e1. doi:10.1016/B978-0-323-67320-4.00069-9

50. Li K, Shrivastava S, Brownley A, Katzel D, Bera J, Nguyen AT, et al. Automated degenerate PCR primer design for high-throughput sequencing improves efficiency of viral sequencing. Virol J. 2012;9: 1–9. doi:10.1186/1743-422X-9-261/FIGURES/3

51. Grubaugh ND, Ladner JT, Kraemer MUG, Dudas G, Tan AL, Gangavarapu K, et al. Genomic epidemiology reveals multiple introductions of Zika virus into the United States. Nature. 2017;546: 401. doi:10.1038/NATURE22400

52. Metsky HC, Matranga CB, Wohl S, Schaffner SF, Freije CA, Winnicki SM, et al. Zika virus evolution and spread in the Americas. Nat 2017 5467658. 2017;546: 411–415. doi:10.1038/nature22402

53. Lin X, Glier M, Kuchinski K, Ross-Van Mierlo T, McVea D, Tyson JR, et al. Assessing Multiplex Tiling PCR Sequencing Approaches for Detecting Genomic Variants of SARS-CoV-2 in Municipal Wastewater. mSystems. 2021;6. doi:10.1128/msystems.01068-21

54. Gohl DM, Garbe J, Grady P, Daniel J, Watson RHB, Auch B, et al. A rapid, cost-effective tailed amplicon method for sequencing SARS-CoV-2. BMC Genomics. 2020;21: 1–10. doi:10.1186/S12864-020-07283-6/FIGURES/4

55. Rebuffo-Scheer C, Bose M, He J, Khaja S, Ulatowski M, Beck ET, et al. Whole Genome Sequencing and Evolutionary Analysis of Human Respiratory Syncytial Virus A and B from Milwaukee, WI 1998-2010. PLoS One. 2011;6. doi:10.1371/JOURNAL.PONE.0025468

56. Kim WK, Kim JA, Song DH, Lee D, Kim YC, Lee SY, et al. Phylogeographic analysis of hemorrhagic fever with renal syndrome patients using multiplex PCR-based next generation sequencing. Sci Reports 2016 61. 2016;6: 1–8. doi:10.1038/srep26017

57. No JS, Kim WK, Cho S, Lee SH, Kim JA, Lee D, et al. Comparison of targeted next-generation sequencing for whole-genome sequencing of Hantaan orthohantavirus in Apodemus agrarius lung tissues. Sci Reports 2019 91. 2019;9: 1–9. doi:10.1038/s41598-019-53043-2

58. Kim WK, No JS, Lee SH, Song DH, Lee D, Kim JA, et al. Multiplex PCR−Based Next-Generation Sequencing and Global Diversity of Seoul Virus in Humans and Rats. Emerg Infect Dis. 2018;24: 249. doi:10.3201/EID2402.171216

59. Watts JM, Dang KK, Gorelick RJ, Leonard CW, Bess JW, Swanstrom R, et al. Architecture and secondary structure of an entire HIV-1 RNA genome. Nature. 2009;460: 711–716. doi:10.1038/nature08237

60. Grubaugh ND, Gangavarapu K, Quick J, Matteson NL, De Jesus JG, Main BJ, et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. Genome Biol. 2019;20: 1–19. doi:10.1186/S13059-018-1618-

7/FIGURES/9

61. Artic Network. 2021 [cited 31 Jan 2024]. Available: https://artic.network/

62. Quick J. nCoV-2019 sequencing protocol v3 (LoCost). In: protocols.io [Internet]. 2020 [cited 12 Mar 2024]. doi:dx.doi.org/10.17504/protocols.io.bp2l6n26rgqe/v3

63. nCoV-2019 sequencing protocol for illumina. In: protocols.io [Internet]. 2021 [cited 12 Mar 2024]. Available: https://www.protocols.io/view/ncov-2019-sequencing-protocol-for-illumina-eq2ly398mgx9/v5

64. Gaudin M, Desnues C. Hybrid Capture-Based Next Generation Sequencing and Its Application to Human Infectious Diseases. Front Microbiol. 2018;9: 27. doi:10.3389/FMICB.2018.02924

65. Sint D, Raso L, Traugott M. Advances in multiplex PCR: Balancing primer efficiencies and improving detection success. Methods Ecol Evol. 2012;3: 898–905. doi:10.1111/j.2041-210X.2012.00215.x

66. Mamanova L, Coffey AJ, Scott CE, Kozarewa I, Turner EH, Kumar A, et al. Target-enrichment strategies for next-generation sequencing. Nat Methods. 2010;7: 111–118. doi:10.1038/nmeth.1419

67. Munyuza C, Ji H, Lee ER. Probe Capture Enrichment Methods for HIV and HCV Genome Sequencing and Drug Resistance Genotyping. Pathogens. 2022;11. doi:10.3390/PATHOGENS11060693

68. Martin S, Heavens D, Lan Y, Horsfield S, Clark MD, Leggett RM. Nanopore adaptive sampling: a tool for enrichment of low abundance species in metagenomic samples. Genome Biol. 2022;23: 1–27. doi:10.1186/S13059-021-02582-X/TABLES/8

69. Yamaguchi J, Olivo A, Laeyendecker O, Forberg K, Ndembi N, Mbanya D, et al.

Universal target capture of HIV sequences from NGS libraries. Front Microbiol. 2018;9: 405367. doi:10.3389/FMICB.2018.02150/BIBTEX

70. Berg MG, Yamaguchi J, Alessandri-Gradt E, Tell RW, Plantier JC, Brennan CA. A Pan-HIV strategy for complete genome sequencing. J Clin Microbiol. 2016;54: 868–882. doi:10.1128/JCM.02479-15

71. Brodin J, Krishnamoorthy M, Athreya G, Fischer W, Hraber P, Gleasner C, et al. A multiple-alignment based primer design algorithm for genetically highly variable DNA targets. BMC Bioinformatics. 2013;14: 1–9. doi:10.1186/1471-2105-14-255/FIGURES/6

72. Yoon H, Leitner T. PrimerDesign-M: a multiple-alignment based multiple-primer design tool for walking across variable genomes. Bioinformatics. 2015;31. doi:10.1093/BIOINFORMATICS/BTU832

73. Ye J, Coulouris G, Zaretskaya I, Cutcutache I, Rozen S, Madden TL. Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. BMC Bioinformatics. 2012;13: 134. doi:10.1186/1471-2105-13-134

74. Zhou S, Hill CS, Spielvogel E, Clark MU, Hudgens MG, Swanstrom R. Unique Molecular Identifiers and Multiplexing Amplicons Maximize the Utility of Deep Sequencing to Critically Assess Population Diversity in RNA Viruses. ACS Infect Dis. 2022;8: 2505–2514. doi:10.1021/ACSINFECDIS.2C00319/SUPPL_FILE/ID2C00319_SI_001.PDF

75. Zhou S, Hill CS, Clark MU, Sheahan TP, Baric R, Swanstrom R. Primer ID Next-Generation Sequencing for the Analysis of a Broad Spectrum Antiviral Induced Transition Mutations and Errors Rates in a Coronavirus Genome. Bio-protocol. 2021;11. doi:10.21769/BIOPROTOC.3938

76. Zhou S, Jones C, Mieczkowski P, Swanstrom R. Primer ID Validates Template Sampling

Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. J Virol. 2015;89: 8540–8555. doi:10.1128/JVI.00522-15/SUPPL_FILE/ZJV999090674SO1.PDF

77. Thompson LR, Sanders JG, McDonald D, Amir A, Ladau J, Locey KJ, et al. A communal catalogue reveals Earth's multiscale microbial diversity. Nat 2017 5517681. 2017;551: 457–463. doi:10.1038/nature24621

78. Public Health Agency of Canada. Update on HIV-1 strain and transmitted drug resistance in Canada: 2012-2013. In: Centre for Communicable Diseases and Infection Control [Internet]. 2017 [cited 31 Jan 2024]. Available: https://www.canada.ca/en/public-health/services/publications/diseases-conditions/hiv-canadian-strain-drug-resistance-surveillance-2012-2013.html#t1

79. Illumina. Nextera XT DNA Library Prep Reference Guide. 2019 [cited 15 Feb 2024]. Available: https://emea.support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/samplepreps_nextera/nextera-xt/nextera-xt-library-prep-reference-guide-15031942-05.pdf

80. Illumina. MiSeq System Denature and Dilute Libraries. 2019 [cited 15 Feb 2024]. Available: www.illumina.com/company/legal.html.

81. Simon Andrews. FastQC: A Quality Control tool for High Throughput Sequence Data. 2010 [cited 8 May 2023]. Available: https://www.bioinformatics.babraham.ac.uk/projects/fastqc/

82. Smith T, Heger A, Sudbery I. UMI-tools: Modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017;27: 491–499. doi:10.1101/GR.209601.116/-/DC1

83. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 2014;30: 2114. doi:10.1093/BIOINFORMATICS/BTU170

84. Langmead B, Salzberg SL. Fast gapped-read alignment with Bowtie 2. Nat Methods. 2012;9: 357. doi:10.1038/NMETH.1923

85. Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, et al. Twelve years of SAMtools and BCFtools. Gigascience. 2021;10: 1–4. doi:10.1093/GIGASCIENCE/GIAB008

86. Wickham H. ggplot2: Elegant Graphics for Data Analysis. Springer-Verlag New York; 2016. Available: https://ggplot2.tidyverse.org

87. Zhou S, Swanstrom R. Fact and Fiction about 1%: Next Generation Sequencing and the Detection of Minor Drug Resistant Variants in HIV-1 Populations with and without Unique Molecular Identifiers. Viruses. 2020;12. doi:10.3390/V12080850

88. World Health Organization. HIV Drug Resistance Report 2021. World Health Organization. 2021. Available: https://www.who.int/publications/i/item/9789240038608

89. Steegen K, van Zyl GU, Claassen M, Khan A, Pillay M, Govender S, et al. Advancing HIV Drug Resistance Technologies and Strategies: Insights from South Africa's Experience and Future Directions for Resource-Limited Settings. Diagnostics 2023, Vol 13, Page 2209. 2023;13: 2209. doi:10.3390/DIAGNOSTICS13132209

90. Bonner K, Siemieniuk RA, Boozary A, Roberts T, Fajardo E, Cohn J. Expanding Access to HIV Viral Load Testing: A Systematic Review of RNA Stability in EDTA Tubes and PPT beyond Current Time and Temperature Thresholds. PLoS One. 2014;9. doi:10.1371/JOURNAL.PONE.0113813

91. Monleau M, Butel C, Delaporte E, Boillot F, Peeters M. Effect of storage conditions of

dried plasma and blood spots on HIV-1 RNA quantification and PCR amplification for drug resistance genotyping. J Antimicrob Chemother. 2010;65: 1562–1566. doi:10.1093/JAC/DKQ205

92. Ávila-Ríos S, Parkin N, Swanstrom R, Paredes R, Shafer R, Ji H, et al. Next-Generation Sequencing for HIV Drug Resistance Testing: Laboratory, Clinical, and Implementation Considerations. Viruses. 2020;12: 617. doi:10.3390/V12060617

93. Lee ER, Parkin N, Jennings C, Brumme CJ, Enns E, Casadellà M, et al. Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing. Sci Rep. 2020;10: 1–10. doi:10.1038/s41598-020-58544-z

94. Shendure J, Balasubramanian S, Church GM, Gilbert W, Rogers J, Schloss JA, et al. DNA sequencing at 40: Past, present and future. Nature. 2017;550: 345–353. doi:10.1038/nature24286

95. Derache A, Iwuji CC, Baisley K, Danaviah S, Marcelin AG, Calvez V, et al. Impact of Next-generation Sequencing Defined Human Immunodeficiency Virus Pretreatment Drug Resistance on Virological Outcomes in the ANRS 12249 Treatment-as-Prevention Trial. Clin Infect Dis An Off Publ Infect Dis Soc Am. 2019;69: 207. doi:10.1093/CID/CIY881

96. Inzaule SC, Hamers RL, Noguera-Julian M, Casadellà M, Parera M, Kityo C, et al. Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. Lancet HIV. 2018;5: e638–e646. doi:10.1016/S2352-3018(18)30177-2

97. Simen BB, Simons JF, Hullsiek KH, Novak RM, MacArthur RD, Baxter JD, et al. Low-Abundance Drug-Resistant Viral Variants in Chronically HIV-Infected, Antiretroviral Treatment–Naive Patients Significantly Impact Treatment Outcomes. J Infect Dis.

2009;199: 693–701. doi:10.1086/596736

98. Ávila-Ríos S, García-Morales C, Matías-Florentino M, Romero-Mora KA, Tapia-Trejo D, Quiroz-Morales VS, et al. Pretreatment HIV-drug resistance in Mexico and its impact on the eff ectiveness of fi rst-line antiretroviral therapy: a nationally representative 2015 WHO survey. Lancet HIV. 2016;3: e579–e591. doi:10.1016/S2352-3018(16)30119-9

99. Gibson RM, Schmotzer CL, Quiñones-Mateu ME. Next-generation sequencing to help monitor patients infected with HIV: Ready for clinical use? Curr Infect Dis Rep. 2014;16: 1–9. doi:10.1007/S11908-014-0401-5/TABLES/1

100. Boyce CL, Beck IA, Styrchak SM, Hardy SR, Wallner JJ, Milne RS, et al. Assessment of minority frequency pretreatment HIV drug-resistant variants in pregnant women and associations with virologic non-suppression at term. PLoS One. 2022;17: e0275254. doi:10.1371/JOURNAL.PONE.0275254

101. Milne RS, Beck IA, Levine M, So I, Andersen N, Deng W, et al. Low-frequency pre-treatment HIV drug resistance: effects on 2-year outcome of first-line efavirenz-based antiretroviral therapy. AIDS. 2022;36: 1949–1958. doi:10.1097/QAD.0000000000003361

102. Maruapula D, Seatla KK, Morerinyane O, Molebatsi K, Giandhari J, De Oliveira T, et al. Low-frequency HIV-1 drug resistance mutations in antiretroviral naïve individuals in Botswana. Med (United States). 2022;101: E29577. doi:10.1097/MD.0000000000029577

103. Moreno A, González C, Góndola J, Chavarría O, Ortiz A, Castillo J, et al. HIV-1 Low-Frequency Variants Identified in Antiretroviral-Naïve Subjects with Virologic Failure after 12 Months of Follow-Up in Panama. Infect Dis Rep. 2023;15: 436–444. doi:10.3390/IDR15040044/S1

104. Ji H, Sandstrom P, Paredes R, Harrigan PR, Brumme CJ, Rios SA, et al. Are We Ready

for NGS HIV Drug Resistance Testing? The Second “Winnipeg Consensus” Symposium. Viruses 2020, Vol 12, Page 586. 2020;12: 586. doi:10.3390/V12060586

105. Lorenz TC. Polymerase Chain Reaction: Basic Protocol Plus Troubleshooting and Optimization Strategies. J Vis Exp. 2012;63: 3998. doi:10.3791/3998

106. Roux KH. Optimization and troubleshooting in PCR. Cold Spring Harb Protoc. 2009;4. doi:10.1101/PDB.IP66

107. Kibirige CN, Manak M, King D, Abel B, Hack H, Wooding D, et al. Development of a sensitive, quantitative assay with broad subtype specificity for detection of total HIV-1 nucleic acids in plasma and PBMC. Sci Reports 2022 121. 2022;12: 1–16. doi:10.1038/s41598-021-03016-1

108. Hebberecht L, Vancoillie L, Schauvliege M, Staelens D, Demecheleer E, Hardy J, et al. Single genome sequencing of near full-length HIV-1 RNA using a limiting dilution approach. J Virol Methods. 2019;274: 113737. doi:10.1016/J.JVIROMET.2019.113737

109. McNulty A, Jennings C, Bennett D, Fitzgibbon J, Bremer JW, Ussery M, et al. Evaluation of dried blood spots for human immunodeficiency virus type 1 drug resistance testing. J Clin Microbiol. 2007;45: 517–521. doi:10.1128/JCM.02016-06

110. Gupta S, Taylor T, Patterson A, Liang B, Bullard J, Sandstrom P, et al. A robust PCR protocol for HIV drug resistance testing on low-level viremia samples. Biomed Res Int. 2017;2017. doi:10.1155/2017/4979252

111. Grossmann S, Nowak P, Neogi U. Subtype-independent near full-length HIV-1 genome sequencing and assembly to be used in large molecular epidemiological studies and clinical management. J Int AIDS Soc. 2015;18: 20035. doi:10.7448/IAS.18.1.20035

112. Zhang G, DeVos J, Medina-Moreno S, Wagar N, Diallo K, Suzanne Beard R, et al.

Utilization of dried blood spot specimens can expedite nationwide surveillance of HIV drug resistance in resource-limited settings. PLoS One. 2018;13. doi:10.1371/JOURNAL.PONE.0203296

113. Ji H, Li Y, Graham M, Liang BB, Pilon R, Tyson S, et al. Next-Generation Sequencing of Dried Blood Spot Specimens: A Novel Approach to HIV Drug-Resistance Surveillance. https://doi.org/103851/IMP1839. 2011;16: 871–878. doi:10.3851/IMP1839

114. Lee SH, Vigliotti VS, Vigliotti JS, Pappu S. Validation of human papillomavirus genotyping by signature DNA sequence analysis. BMC Clin Pathol. 2009;9: 3. doi:10.1186/1472-6890-9-3

115. Amougou-Atsama M, Jean Adrien Atangana P, Noah Noah D, Fewou Moundipa P, Pineau P, Njouom R. The role of hepatitis C virus genotypes and core mutations in hepatocellular carcinoma in Cameroon. J Viral Hepat. 2020;27: 880–885. doi:10.1111/JVH.13303

116. Aralaguppe SG, Siddik AB, Manickam A, Ambikan AT, Kumar MM, Fernandes SJ, et al. Multiplexed next-generation sequencing and de novo assembly to obtain near full-length HIV-1 genome from plasma virus. J Virol Methods. 2016;236: 98–104. doi:10.1016/j.jviromet.2016.07.010

117. Ode H, Matsuda M, Matsuoka K, Hachiya A, Hattori J, Kito Y, et al. Quasispecies Analyses of the HIV-1 Near-full-length Genome With Illumina MiSeq. Front Microbiol. 2015;6: 1258. doi:10.3389/FMICB.2015.01258

118. Ryou S, Yoo M, Kim K, Kim S, Kim S Il, Kim YJ, et al. Characterization of HIV-1 recombinant and subtype B near full-length genome among men who have sex with men in South Korea. Sci Reports 2021 111. 2021;11: 1–9. doi:10.1038/s41598-021-82872-3

119. Hull NC, Yeboah E, Tsaknaridis LJ, Makris V. An NGS amplicon tiling protocol for HIV-1 drug resistance detection using Illumina® COVIDSeq™ Assay Kit. In: protocols.io [Internet]. 2023 [cited 26 Jan 2024]. Available: https://dx.doi.org/10.17504/protocols.io.n92ldmq4ol5b/v2

120. Botha JC, Byott M, Spyer MJ, Grant PR, Gärtner K, Chen WX, et al. Sensitive HIV-1 DNA Pol Next-Generation Sequencing for the Characterisation of Archived Antiretroviral Drug Resistance. Viruses. 2023;15: 1811. doi:10.3390/v15091811

121. Chandra V, Li L, Le Roux O, Zhang Y, Howell RM, Rupani DN, et al. Gut epithelial Interleukin-17 receptor A signaling can modulate distant tumors growth through microbial regulation. Cancer Cell. 2024;42: 85-100.e6. doi:10.1016/J.CCELL.2023.12.006

122. Hossain AA, Pigli YZ, Baca CF, Heissel S, Thomas A, Libis VK, et al. DNA glycosylases provide antiviral defence in prokaryotes. Nat 2024. 2024; 1–7. doi:10.1038/S41586-024-07329-9

123. Lim JJ, Diener C, Wilson J, Valenzuela JJ, Baliga NS, Gibbons SM. Growth phase estimation for abundant bacterial populations sampled longitudinally from human stool metagenomes. Nat Commun 2023 141. 2023;14: 1–13. doi:10.1038/S41467-023-41424-1

124. MacConaill LE, Burns RT, Nag A, Coleman HA, Slevin MK, Giorda K, et al. Unique, dual-indexed sequencing adapters with UMIs effectively eliminate index cross-talk and significantly improve sensitivity of massively parallel sequencing. BMC Genomics. 2018;19: 1–10. doi:10.1186/S12864-017-4428-5/FIGURES/5

125. Manyana S, Pillay M, Gounder L, Khan A, Moodley P, Naidoo K, et al. Affordable drug resistance genotyping of HIV-1 reverse transcriptase, protease and integrase genes, for resource limited settings. AIDS Res Ther. 2023;20: 1–12. doi:10.1186/S12981-023-

00505-3/TABLES/5

126. Sun J, Philpott M, Loi D, Li S, Monteagudo-Mesas P, Hoffman G, et al. Correcting PCR amplification errors in unique molecular identifiers to generate accurate numbers of sequencing molecules. Nat Methods. 2024; 1–5. doi:10.1038/s41592-024-02168-y

127. Keys JR, Zhou S, Anderson JA, Eron JJ, Rackoff LA, Jabara C, et al. Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain. AIDS Res Hum Retroviruses. 2015;31: 658. doi:10.1089/AID.2014.0031

128. Faraci G, Park SY, Dubé MP, Lee HY. Full-spectrum HIV drug resistance mutation detection by high-resolution complete pol gene sequencing. J Clin Virol. 2023;164: 105491. doi:10.1016/J.JCV.2023.105491

129. Miyazato P, Katsuya H, Fukuda A, Uchiyama Y, Matsuo M, Tokunaga M, et al. Application of targeted enrichment to next-generation sequencing of retroviruses integrated into the host human genome. Sci Rep. 2016;6. doi:10.1038/SREP28324

130. MacKenzie M, Argyropoulos C. An Introduction to Nanopore Sequencing: Past, Present, and Future Considerations. Micromachines. 2023;14. doi:10.3390/MI14020459

131. Minei R, Hoshina R, Ogura A. De novo assembly of middle-sized genome using MinION and Illumina sequencers 06 Biological Sciences 0604 Genetics. BMC Genomics. 2018;19: 1–12. doi:10.1186/S12864-018-5067-1/FIGURES/6

132. Ashton PM, Nair S, Dallman T, Rubino S, Rabsch W, Mwaigwisya S, et al. MinION nanopore sequencing identifies the position and structure of a bacterial antibiotic resistance island. Nat Biotechnol 2014 333. 2014;33: 296–300. doi:10.1038/NBT.3103

133. Oxford Nanopore Technologies. Nanopore Sequencing Accuracy. 2023 [cited 21 Apr

2024]. Available: https://nanoporetech.com/platform/accuracy

134. Quick J, Loman NJ, Duraffour S, Simpson JT, Severi E, Cowley L, et al. Real-time, portable genome sequencing for Ebola surveillance. Nature. 2016;530: 228–232. doi:10.1038/nature16996

135. Wang J, Moore NE, Deng YM, Eccles DA, Hall RJ. MinION nanopore sequencing of an influenza genome. Front Microbiol. 2015;6. doi:10.3389/FMICB.2015.00766

136. Charre C, Ginevra C, Sabatier M, Regue H, Destras G, Brun S, et al. Evaluation of NGS-based approaches for SARS-CoV-2 whole genome characterisation. Virus Evol. 2020;6: 75. doi:10.1093/VE/VEAA075

137. Mori M, Ode H, Kubota M, Nakata Y, Kasahara T, Shigemi U, et al. Nanopore Sequencing for Characterization of HIV-1 Recombinant Forms. Microbiol Spectr. 2022;10. doi:10.1128/SPECTRUM.01507-22

138. Ng TTL, Su J, Lao HY, Lui WW, Chan CTM, Leung AWS, et al. Long-Read Sequencing with Hierarchical Clustering for Antiretroviral Resistance Profiling of Mixed Human Immunodeficiency Virus Quasispecies. Clin Chem. 2023;69: 1174–1185. doi:10.1093/clinchem/hvad108

139. Lambrechts L, Bonine N, Verstraeten R, Pardons M, Noppe Y, Rutsaert S, et al. HIV-PULSE: a long-read sequencing assay for high-throughput near full-length HIV-1 proviral genome characterization. Nucleic Acids Res. 2023;51: e102. doi:10.1093/NAR/GKAD790

140. Malet I, Subra F, Charpentier C, Collin G, Descamps D, Calvez V, et al. Mutations Located outside the Integrase Gene Can Confer Resistance to HIV-1 Integrase Strand Transfer Inhibitors. MBio. 2017;8. doi:10.1128/MBIO.00922-17

141. Fun A, Wensing AMJ, Verheyen J, Nijhuis M. Human Immunodeficiency Virus gag and protease: partners in resistance. Retrovirology. 2012;9: 63. doi:10.1186/1742-4690-9-63