DROPLET DIGITAL PCR DETECTION AND NEXT-GENERATION SEQUENCING OF HIV-1 IN BREASTMILK OF BREASTFEEDING WOMEN ON ANTIRETROVIRAL TREATMENT

By

Jennah Nivashni Hurree

HRRJEN011

Supervisor: Dr. Melissa-Rose Abrahams

Supervisor: Dr. Marvin Hsiao

Thesis submitted to the University of Cape Town in fulfilment of the degree;

MSc (Med) in Medical Virology

Division of Medical Virology

Department of Pathology Faculty

of Health Sciences

University of Cape Town

Submission Date: 28 June 2024

Declaration

I, Jennah Nivashni Hurree, hereby declare that the work on which this dissertation/thesis is based is my original work (except where acknowledgements indicate otherwise) and that neither the whole work nor any part of it has been, is being, or is to be submitted for another degree in this or any other university.

I empower the university to reproduce for the purpose of research either the whole or any portion of the contents in any manner whatsoever.

Signature: Signed by candidate

Date: 20.06.24

Table of contents

Acknowledgements

I would like to express my appreciation to those who have contributed to this degree.

I am profoundly grateful for my supervisor Dr. Melissa-Rose Abrahams for her endless patience, inspiration, gentle push, dedication and for going the extra mile to ensure that the best possible version of my work is produced. I also extend my gratitude to my supervisor Dr. Marvin Hsiao for his valuable support and feedback.

I am equally thankful to Ms. Lynn Tyers and Dr. Sherazaan Ismail for their reliability and willingness to help, alongside Dr. Abrahams, who have collectively been the backbone of my laboratory work. A special thanks to Shorok Sebaa and Josephine Peka for their kindness and eagerness to assist and to Dr. Chivonne Moodley for her encouragement and meticulous attention to detail.

To the DolPHIN study team, I would like to thank Nai-Chung Hu, Catriona Waitt, Helen Reynolds, Saye Khoo and Landon Myer for the provision of breastmilk cell samples.

To the University of Cape Town and the Queen Elizabeth College for encouraging me to dream big and for providing me with the resources to so. I also have a special thought for my friends Ruth, Nkosazana, Hansini, Sheetul and Urvashi for their support and cheer as we pursue our individual aspirations.

Words do not suffice to express my gratitude for my parents, Sooriabye and Vidianand Hurree who have always encouraged me to aim for the moon, believe in my potential and provide unwavering support towards my endeavours and growth.

I am immensely thankful for the support of my family and wish to acknowledge their special contributions: Ruby and Raj Seetaram, Tara and Shriram Malloo for their unconditional care and guidance, the Periatamby family, including Vassanda who has been my inspiration, Vanesha and Vishal Ramdin, Kean Hurree and family for their unparalleled support, Geeta and Ajay Bappoo as well as the Poorun family who never fail to see me off at the airport.

Last but not least, I am deeply grateful to Duncan Maneveld and family for being my home away from home.

List of Abbreviations

AgSCs	Antigen secreting cells
AIDS	Acquired Immunodeficiency Syndrome
ART	Antiretroviral therapy
ARV	Antiretroviral
BAN	Breastfeeding, antiretrovirals and nutrition
bp	Base pairs
CAF	Central Analytical Facility
CAPRISA	Centre for the AIDS Programme of Research in South Africa
cART	Combination ART
CAV	Cell-associated virus
CFV	Cell-free virus
Cq	Quantification cycle
DC-SIGN	Dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin
ddPCR	Droplet digital PCR
DNA	Deoxyribonucleic acid
dNTPs	Deoxynucleotide triphosphate
DolPHIN-2	Dolutegravir in pregnant HIV infected mothers and their neonates
DTG	Dolutegravir
EFV	Efavirenz

e-MTCT	Elimination of mother to child transmission
env	Envelope gene
FACS	Fluorescence-Activated Cell Sorting
GALT	Gut-associated lymphoid tissue
gDNA	Genomic DNA
GIT	Gastrointestinal tract
HAART	Highly active antiretroviral therapy
HIV	Human immunodeficiency virus
IDT	Integrated DNA Technologies
IL-8	Interleukin-8
IMPAACT PROMISE	International Maternal Paediatric Adolescent AIDS Clinical Trials Group - Promoting Maternal Infant Survival Everywhere
int	Integrase gene
iNVP	Infant Nevirapine
KiBS	Kisumu Breastfeeding Study
LANL	Los Alamos National Laboratory
LAV	Lymphadenopathy-Associated Virus
LBW	Low birth weight
LTR	Long terminal repeat
M cells	Microfold cells
MACS	Magnetic-activated cell sorting
mART	Maternal ART
MEC	Mammary epithelial cells

mgH_20	Molecular grade water
ml	Millilitre
mRNA	Messenger RNA
MTCT	Mother to child transmission
NCBI	National Center for Biotechnology Information
NGS	Next generation sequencing
NRTI	Nucleoside reverse-transcriptase inhibitor
ng	Nanogram
nm	Nanometer
nM	Nanomolar
NTC	Non-template controls
PACTG	Paediatric AIDS Clinical Trials Group
PAL	Pooled amplicon library
PBMCs	Peripheral blood mononuclear cells
PCR	Polymerase chain reaction
PLWH	People living with HIV
PMTCT	Prevention of mother to child transmission
PrEP	Pre-exposure prophylaxis
PTB	Preterm birth
qPCR	Quantitative PCR
RFU	Relative fluorescence units
RNA	Ribonucleic acid
rpm	Revolutions per minute

RPMI-1640	Roswell Park Memorial Institute
R1	Round 1
R2	Round 2
SCM	Subclinical mastitis
SOP	Standard operating procedure
ssDNA	Single strand DNA
TBE	Tris-Borate EDTA
vRNA	Viral genomic RNA
WHO	World Health Organization
µl	Microlitre
µg	Microgram

List of Tables

List of Figures

Abstract

Persistence of HIV-1 in a stable latent reservoir in breastmilk has been associated with low-level mother to child transmission despite maternal antiretroviral therapy (ART). We aimed to evaluate and optimize the sensitivity of HIV-1 DNA quantitation in breastmilk using droplet digital PCR (ddPCR) and sequencing of viral variants in breastmilk using Illumina MiSeq next generation sequencing to inform transmission risk and efficacy of maternal ART clinical trials.

HIV-negative breastmilk samples were spiked with 8E5 cells (a cell line where each cell contains one replication-defective HIV genome) at concentrations ranging from 10 000 - 1 8E5 cell(s)/million. Genomic DNA was extracted with the QIAGEN AllPrep DNA/RNA Mini Kit (QIAGEN, Hilden, Denmark) and ddPCR was performed using the ddPCR Supermix for Probes (No dUTP) (BIO-RAD, California, USA), HIV-1 *gag* and/or *pol* and housekeeping gene *rpp30* primers and probes, BanII restriction enzyme, CutSmart buffer (New England Biolabs, Massachusetts, USA) and the QX200^{TM} ddPCR System (BIO-RAD, California, USA). The envelope gene V1-V2 region was amplified using the KAPA2G Fast Multiplex Mix (Roche, Basel, Switzerland) and Expand High Fidelity PCR enzyme (Roche diagnostics, Basel Switzerland) with Nextera DNA CD indexes and sequenced via the Illumina Miseq v3 platform. Genomic DNA recovery was low (ranging from 25.12 - 106.85 ng/µl with an average of 67.18 ng/µl) when spiking in a background of 1 million cells, resulting in HIV DNA being detectable by ddPCR only in samples spiked with the highest 8E5 cell concentration. When repeating this experiment with 8E5 cell spike-in concentrations ranging from 50 000 - 5 8E5 cells/million in a total of 5 million cells, genomic DNA recovery was higher (ranging from 498.99 - 757.28 ng/µl with an average of 631.32 ng/µl) resulting in a detection sensitivity of 44.3 *gag* copies/million cells with a lowest per reaction HIV copy number detection of 7 *gag* copies. We noted that 8E5 cell line HIV genome copies/cell varied across different passages, ranging from 0.4 - 0.8 copies/cell instead of the expected 1 copy/cell.

HIV-negative breastmilk samples spiked with 8E5 cells ranging from 50 000 - 50 8E5 cells/million were likewise successfully PCR amplified and sequenced (gDNA input at 1000 ng), with sequences matching the 8E5 HIV-1 /LAV genome, as expected. Additionally, this method

was successfully applied to genomic DNA from PBMCs of women on ART from the CAPRISA 002 cohort, Kwa-Zulu Natal, confirming the sensitivity of this method to amplify low-copy number HIV templates. ddPCR and Illumina MiSeq methods were applied to cells from 2 ml breastmilk samples from three women with blood viral loads of 20 - 331 copies/ml from the Dolutegravir in pregnant HIV infected mothers and their neonates (DolPHIN-2) trial, yet no HIV genome copies above background level were detected and no HIV targets could be PCR amplified.

Viral quantification and sequencing from breastmilk is difficult particularly when viral load is low in the presence of ART and no guidelines have been established for processing breastmilk. Both ddPCR and Illumina Miseq were successful at quantifying and sequencing low levels ($1.3\log_{10}$) of HIV from spiked-in breastmilk samples. HIV copies below background levels obtained from breastmilk samples of the DolPHIN-2 study potentially demonstrate effective viral suppression in breastmilk by integrase inhibitor Dolutegravir-based ART. However, the number of samples analysed was low and cell pellets available for analysis possibly contained too few cells for ddPCR detection and Illumina Miseq of HIV, based on our observations from the spike-in experiments.

Chapter 1 : Literature Review

1.1 Introduction

Mother to child transmission (MTCT) of the human immunodeficiency virus (HIV) can take place at several time points: *in utero* (transplacental passage), intrapartum (during birthing), and postpartum through breastfeeding (Nakamura *et al.*, 2017). MTCT rates range from 15% to 45% in the absence of preventive measures with estimated transmission rates of 20% to 25% occurring in utero, 35% to 50% during the intrapartum period and 25% to 45% through breastfeeding (Amin *et al.*, 2021; Luzuriaga and Mofenson, 2016).

Breastmilk provides important nutrients, bioactive compounds and protective immune cells to promote infant growth and immune development (Lyons *et al.*, 2020). Breastmilk also protects infants against a wide range of illnesses such as gastrointestinal and respiratory infections, chronic diseases and sudden infant death syndrome (Feldman-Winter, 2012). While formula is a safe alternative to breastmilk in the case of maternal HIV infection (Read, 2003), it has been shown that in the context of antiretroviral therapy (ART), avoidance of breastfeeding leads to worse outcomes in resource limited settings (Doherty *et al.*, 2011). Therefore, the World Health Organization (WHO) recommends exclusive breastfeeding for the first 6 months of life followed by complementary food with sustained breastfeeding for at least 12 months and up to 24 months or longer under strict ART adherence (World Health Organisation [WHO], 2023). Furthermore, the WHO recommends that breastfed infants at high risk of acquiring HIV receive prophylaxis with Zidovudine and Nevirapine or Nevirapine only for the first 12 weeks of life and that breastfed infants with mothers on ART receive 6 weeks of daily Nevirapine (WHO, 2016).

According to UNAIDS reports from 2022, the rate of vertical transmission was 11% worldwide and 3% in South Africa. With 98% of pregnant women receiving ART in 2022, South Africa is considered a global leader in the prevention of mother to child transmission (PMTCT). In the same year, 4.1 million HIV exposed children remained uninfected, thus avoiding 47 000 new infections due to PMTCT (UNAIDS, no date). Without any intervention, the rate of MTCT ranged from 15 – 20 % and 20 – 45 % among non-breastfed and breastfed infants respectively. The latter varied according to the duration of breastfeeding (Njom Nlend, 2022; Tolossa *et al.*, 2020). In contrast, the WHO initiative, elimination of mother to child transmission (e-MTCT) programme has successfully reduced MTCT rates to 5% and 2% in breastfeeding and

non-breastfeeding populations respectively in South Africa (Ishikawa *et al.*, 2016; Itiola, Goga and Ramokolo, 2019)

Despite ART reducing maternal viral load to undetectable levels, some cases of MTCT are still recorded, thus raising the concern of viral reservoirs in breastmilk (Njom Nlend, 2022; Valea *et al.*, 2011; Van De Perre *et al.*, 2012). Following infection, HIV-1 integrates its genetic material into the genome of CD4+ T cells, forming latent proviral reservoirs in various anatomical sites within the host's body. Studies have showed that in individuals on ART, this included breastmilk (Lehman *et al.*, 2008; Valea *et al.*, 2011). Highly active antiretroviral therapy (HAART) does not impact the reservoir, resulting in residual HIV-1 transmission via breastfeeding (Lehman *et al.*, 2008). To date, most of the research on HIV persistence has centred on blood samples, leading to the development of numerous techniques for quantification and genetic characterisation of viruses in this particular environment. Conversely, there is a lack of comprehensive data about other anatomical compartments where the virus can persist at very low levels.

Sensitive methods that facilitate quantification and characterization of persistent low-level HIV in breastmilk may offer valuable insights about the risk of MTCT during breastfeeding in the presence of ART, as well as inform strategies to improve maternal prophylaxis. This review will explore HIV vertical transmission with a focus on transmission through breastmilk. It will broadly discuss the HIV life cycle and latent reservoir formation before bringing particular attention to the cells forming a viral reservoir in breastmilk, the different forms of HIV contributing to MTCT through breastfeeding (cell-associated virus (CAV) and cell-free virus (CFV)), mechanisms of MTCT, and other risk factors contributing to transmission. PMTCT will also be reviewed briefly. Lastly, methods to quantify and sequence HIV-1 deoxyribonucleic acid (DNA) in individuals on ART will be elaborated upon.

1.2 The HIV life cycle

HIV is a lentivirus from the family *Retroviridae* that is renowned for challenging the central dogma of biology. These viruses have two identical copies of single-stranded Ribonucleic

acid (RNA) genomes which get reverse transcribed, form a DNA duplex and integrate into the host's DNA to produce a proviral genome. Messenger RNA (mRNA) and viral genomic RNA (vRNA) from the integrated genome get transcribed, leading to continuous viral propagation and lifelong persistent infections. Retroviral genomes include three main genes following a 5'-gag-pol-env-3' structure which translate proteins enabling viral replication (Meissner, Talledge and Mansky, 2022).

The first step in the HIV life cycle is binding and entry (**Figure 1.1**). Viral heterodimer proteins gp120 and gp41 form a trimeric complex on the viral envelope, which facilitates viral adhesion and entry into target cells. Hence, gp120 binds the CD4+ receptor and CCR5 or CXCR4 chemokine co-receptors aided by gp41 which facilitates fusion of the viral and host cell membranes. The second step, called uncoating, leads to the release of viral RNA into the host cytoplasm. Third, Reverse Transcriptase converts viral RNA into double-stranded DNA. Fourth, the Integrase enzyme cleaves the 3' ends of the newly formed DNA helix to create sticky ends which facilitates proviral integration into the host genome after entry into the nucleus (**Figure 1.1**, green shaded region). Fifth, viral protein synthesis and assembly occurs. The provirus may be actively transcribed to mRNA or may remain transcriptionally silent (latent), sometimes for years (Re *et al.*, 2010; Tavassoli, 2011). In activated cells, the mRNA migrates to the cytoplasm for the synthesis of viral regulatory, structural and enzymatic proteins to form mature viral particles whereby two viral RNA strands link with replicative enzymes while core proteins join around them to form a capsid. The immature virion moves to the host cell surface where they acquire a new envelope and lastly, budding occurs and virions are released into the extracellular space (Chen, 2019; Fanales-Belasio *et al.*, 2010; Ramdas *et al.*, 2020).

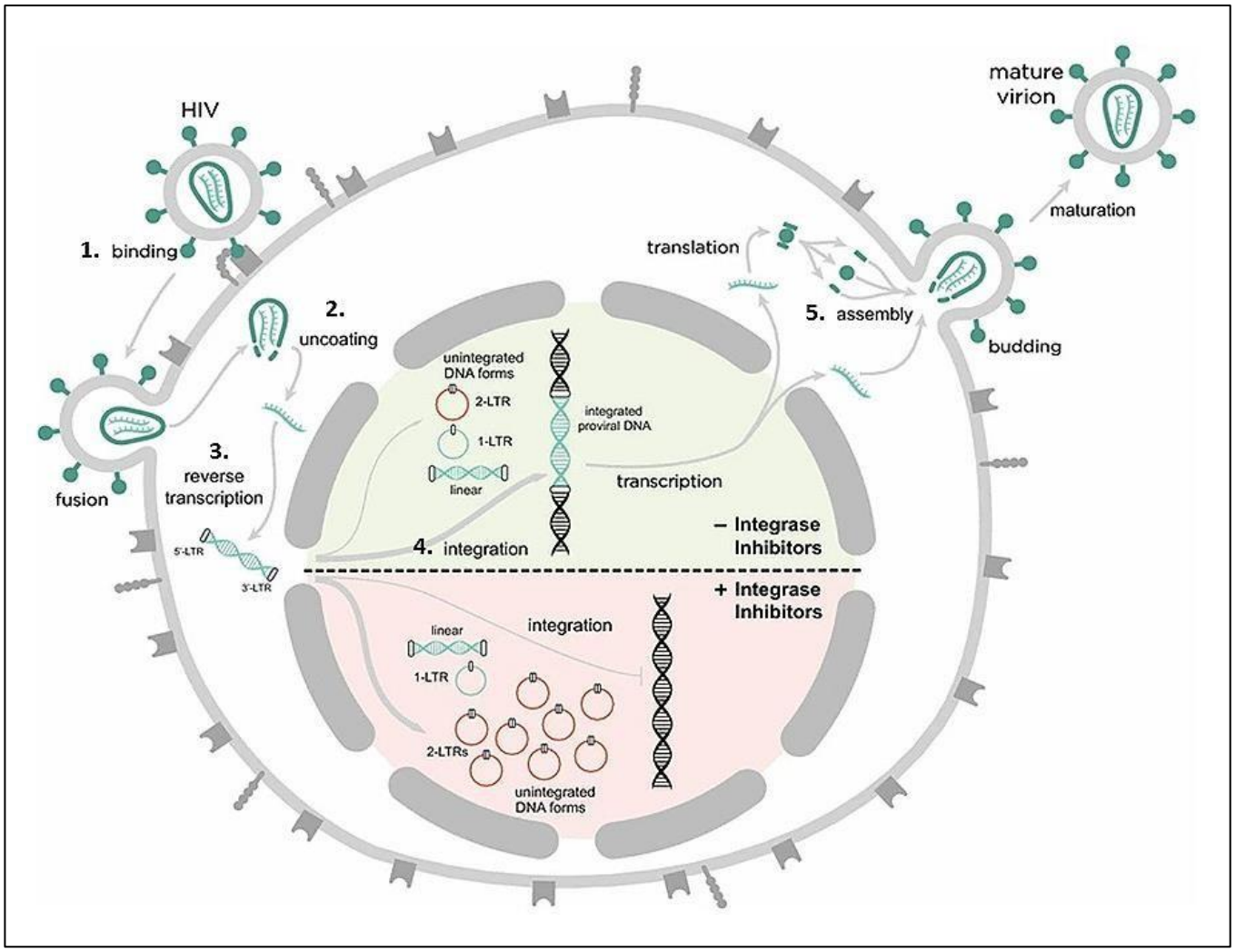

Figure 1.1. HIV-1 life cycle with and without integrase inhibitors illustrating attachment of a virion to the host membrane receptors, fusion, uncoating and release of viral genome in the host cytoplasm followed by reverse transcription and integration of viral DNA into the host genome. Proviral DNA is translated to produce viral proteins which form new virions and bud off. On the other hand, unintegrated viral DNA prevails both in linear and circular forms (1-LTR and 2-LTR circles) (Martinez-Picado *et al.*, 2018).

About 10 – 30 % of viral cDNA synthesized gets integrated into the host genome. The majority of unintegrated cDNA exists in linear and circular forms with relative abundance highest for linear unintegrated DNA followed by integrated provirus, 1-long terminal repeat (1-LTR circles) (~ 10%), and lastly 2-LTR circles (~ 1%) (Vandegraaff *et al.*, 2001). LTR circles, also

called episomal DNA, are closed circular DNAs which cannot replicate autonomously. Nonetheless, it has been suggested that unintegrated HIV DNA may be transiently transcribed and allow virus production, latency and immune responses, particularly in the case of direct infection of resting CD4+ T cells (Chan *et al.*, 2016). Linear unintegrated DNA are precursors to the provirus and may get tagged for apoptosis due to their free ends resembling a double strand break in the chromosome. On the other hand, circularising DNA as a repair mechanism may prevent apoptotic signals (Li, 2001). Whether an accumulation of unintegrated viral DNA may take part in pathogenesis remains uncertain (Bergeron and Sodroski, 1992; Martinez- Picado *et al.*, 2018).

ART effectively supresses viral load but despite sustained undetectable plasma viral RNA levels, episomal HIV DNA, mainly 2-LTR circles, have been detected in many people living with HIV (PLWH) on ART (Pace, Graf and O'Doherty, 2013). Indeed, the use of integrase inhibitors has shown an increase in the number of 2-LTR circles (**Figure 1.1**, pink shaded region) (Hatano *et al.*, 2013).

1.3 HIV latent reservoir formation

Despite the efficacy of ART to reduce HIV viral load to undetectable levels, treatment is lifelong and viral rebound occurs within weeks of treatment interruption (Chun *et al.*, 1995; Chun *et al.*, 1997). This is due to the presence of a stable latent HIV reservoir mainly in resting CD4+ T cells which remains unaffected by ART (Chun *et al.*, 1997; Finzi *et al.*, 1997). Most resting CD4+ T cells are usually non-permissive for active HIV-1 replication and lack CCR5. Resting CD4+ T cells also express very low levels of deoxynucleotide triphosphate (dNTPs), thus preventing reverse transcription of viral RNA into DNA within the host cell. In case of completion of reverse transcription, viral genome integration into the host DNA follows. Nonetheless, a lack of activated host factors needed for HIV-1 transcription and elongation of transcripts prevents viral gene expression. Since resting CD4+ T cells exhibit so many obstacles to viral replication, establishment of the latent reservoir likely occurs by the infection of activated CD4+ T cells returning to a resting memory state (Chun *et al.*, 1995). This is consistent with other studies that found that the latent reservoir consists mainly of resting

memory CD4+ T cells (Chen *et al.*, 2022). Moreover, it was shown that resting CD4+ T-cells with central memory phenotype were the major source of replication-competent HIV in PLWH on combination ART (cART) (Finzi *et al.*, 1997).

A complex set of epigenetic modifications further silence the provirus, resulting in a stable, integrated, transcriptionally silent provirus in a long-lived memory T cell. Following reservoir establishment, proviral transcription levels are very low, producing negligible virus and viral proteins. Since the latently infected cells are not recognised by the immune system, they remain unaffected by cytotoxic events, thus existing stably for a long time and allowing HIV persistence (Chen *et al.*, 2022; García *et al.*, 2018).

Following infection, many reservoir cells have demonstrated viral proliferation whereby the integrated provirus (intact (preserved genome and replication competent) and defective (with large deletions and hypermutations (Ho *et al.*, 2013))) is copied into each daughter cell with minimal error (Siliciano and Siliciano, 2022). Integration of the provirus may lead to a promoter-insertion type of mutagenesis which changes a relevant host gene expression to favour survival and/or proliferation of the infected cells (Siliciano and Siliciano, 2022). Studies have found the presence of massive clonal expansion with cells harbouring inducible, replication-competent proviruses, thus resulting in residual viremia (García *et al.*, 2018). This cell proliferation is thought to maintain most of infected cells during ART and help structure the constitution and location of the proviral population (Cohn, Chomont and Deeks, 2020).

Contrary to perception, the latent reservoir is in fact dynamic and continuously changes in composition while maintaining a relatively constant total size, with a half-life of 44 months (Siliciano *et al.*, 2003). While latency is established early (a few days) after initial infection (Whitney *et al.*, 2014) and early cART initiation can reduce reservoir size (Ananworanich *et al.*, 2016), it does not prevent establishment of the latently infected CD4+ T cell pool. Viral rebound may occur within days of ART discontinuation, suggesting early and continuous reservoir formation even before therapy initiation (García *et al.*, 2018). Interestingly, more recent studies have found that most viruses enter the reservoir around the time that ART is started (Abrahams *et al.,* 2019; Brodin *et al.*, 2016; Pankau *et al.*, 2020). ART initiation may therefore alter the host environment and suppress viral replication but in turn encourage establishment of long-lived cells, some of which may be latently infected with replication-competent HIV-1. Indeed, Abrahams *et al.,* suggested that most of the viral reservoir stabilizes

during the period concurrent with CD4+ T cell count increase post-ART, accompanied by fast reduction in viral load as well as immune activation markers. Hence, they proposed that ART initiation may promote short lived HIV-infected cells to become long lived memory T cells with enhanced proliferative capacity (Abrahams *et al.,* 2019).

With less than 2% of total body lymphocytes residing in peripheral blood, several anatomical tissue sites serve as viral reservoirs where mechanisms such as cell-to-cell transmission may enhance HIV-1 replication. Such sites include the gastrointestinal tract, the liver, the central nervous system, the lymph nodes, the genital tract and the breast tissue (Svicher *et al.*, 2014).

1.4 HIV mother to child transmission via breastmilk

Breastmilk accounts for 30% to 50% of all global HIV MTCTs, with an estimated transmission rate of 0.74% per month of breastfeeding (Prendergast *et al.*, 2019). HIV transmission through breastfeeding is linked to the viral load in breastmilk (Richardson *et al.*, 2003; Salazar-Gonzalez *et al.*, 2011). Breastfeeding has been linked to both CFV and CAV forms of HIV-1, however research suggests that CAV may be the primary contributor of MTCT (Shapiro *et al.*, 2005). In contrast to CFV, antiretroviral (ARV) regimens do not yield a significant reduction in CAV load within breastmilk (Shapiro *et al.*, 2005).

Therefore, complete avoidance of breastfeeding is the most effective way to prevent HIV-1 vertical transmission though breastmilk (Read, 2003). A small study in South Africa reported a 15% increased risk of MTCT among breastfed as compared to formula fed infants (Bobat *et al.,* 1997). Furthermore, longer periods of breastfeeding have been linked to a higher risk of MTCT. Indeed, a meta-analysis of published data from prospective cohort studies of HIV-1-infected women and their children reported that HIV-1 MTCT by breastmilk occurred in 21% of women who breastfed for 3 months or longer compared 13% of those who breastfed for less than 2 months (John *et al.*, 2001). Additionally, mixed feeding is associated with an increased risk of HIV-1 transmission and infant mortality. For instance, a study carried out in Cameroon reported HIV-1 MTCT rates of 20% among mixed fed infants in contrast to 2.2% among exclusively breastfed infants and 3.77% among exclusively replacement fed infants, all born to HIV positive women. The mortality rate of mixed fed infants was also reported to

be 5.4 and 5.5 - fold higher than exclusively breastfed infants and exclusively replacement fed infants, respectively (Njom Nlend *et al.*, 2018).

While refraining from breastfeeding may be most effective in preventing MTCT, it is impractical in resource-constrained regions such as sub-Saharan Africa. Factors such as, high expense of infant formula, stigma associated with replacement feeding, lack of protective qualities of breastmilk, and unavailability of clean water render complete avoidance of breastfeeding challenging (Gantt *et al.*, 2010; Koulinska *et al.*, 2006).

1.4.1 Mechanisms of HIV transmission by breastfeeding

In spite of the protective effects of polarized epithelial barriers lining the mucosa (Milligan and Overbaugh, 2014), most MTCT events via breastfeeding are thought to occur across infant oral and/or buccal mucosa and the upper gastrointestinal tract (GIT) mucosal surfaces which come in contact with HIV-1 in breastmilk, thus serving as the main ports of entry for viral transmission (Dorosko and Connor, 2010). HIV may infect susceptible lymphocytes in the epithelium and/or reach other cells by crossing the barrier. Additionally, the presence of many HIV susceptible target cells (CD4+ CCR5+ T cells) have been reported in foetal and infant gut epithelia (Bunders *et al.*, 2012), suggesting possible infection sites in the event of surface contact with HIV-infected fluids. Alternatively, breaks in the mucosa could facilitate viral entry into the epithelium (Milligan and Overbaugh, 2014).

Direct HIV MTCT during breastfeeding may also occur via HIV infected blood from cracked nipples through abrasions in the mucous membranes of the infant's mouth (Valea *et al.*, 2011). Transcytosis is another possible mechanism by which HIV may be transported across cells. In infants, transcytosis may transfer HIV from the GIT lumen for release into basolateral epithelium cells, thus bringing the virus in touch with susceptible cells for infection (Milligan and Overbaugh, 2014). Tonsils may also act as entry for HIV transmission via transcytosis. The latter consist of microfold cells (M cells) capable of HIV-1 replication in proximity to lymphocytes and dendritic cells (John-Stewart *et al.*, 2004). It is also possible that infant intestinal mucosae actively traffic HIV-1 from the lumen to the submucosa via M cells or enterocytes. Hence, cell associated viral particles may penetrate the mucosa into the submucosa through mucosal breaches or via transcytosis (Valea *et al.*, 2011).

On average, an infant ingests about 178 spontaneous HIV-1 secreting cells from breastmilk per day the first four months postpartum. Since an actively infected cell produces at least 1000 virions (Chun *et al.*, 1997; Dimitrov *et al.*, 1993), an infant may be exposed to about 178 000 cell-associated HIV-1 particles on a daily basis. Therefore, despite undetectable CFV in breastmilk, infant mucosae may be exposed to high levels of viral particles released by infected CD4+ T cells (Valea *et al.*, 2011).

1.4.2 Compartmentalisation and the origin of breastmilk viruses

Defining the origin and genetic composition of breastmilk HIV is essential to develop strategies for the PMTCT and past investigations have revealed controversial data surrounding compartmentalization of HIV variants in breastmilk. Some studies have found evidence of compartmentalisation between plasma and breastmilk as revealed by distinct patterns of viral populations between blood and breastmilk of mothers living with HIV (Becquart *et al.*, 2002). Moreover, Andreotti *et al*. reported higher overall median genetic distances between breastmilk CAV and breastmilk CFV as compared to that between plasma viruses and breastmilk CFV. The genetic distance between plasma viruses and breastmilk CFV as well as between breastmilk CFV and breastmilk CAV were higher in the HAART group as compared to the ART-naïve group. They also observed positive correlation between plasma HIV-RNA and breastmilk CFV. These observations indicate that part of breastmilk viruses originate from plasma although breastmilk viral populations differ from that of plasma and may replicate under selective drug pressure (Andreotti *et al.*, 2009).

Another observation, by Becquart *et al.* was that the main proviral variant in breastmilk corresponded to a minor variant in plasma in 67% of mothers, while in 33% of mothers, none of the major breastmilk variants were represented in the HIV-1 blood population. These observations suggested that infection of breastmilk cells either occurred with a selected set of HIV-1 variants migrating from the blood compartment to the breast tissue or from local mucosal viral replication when breastmilk proviruses forming part of the latent reservoir become replicative during pregnancy to contribute to MTCT by breastfeeding. Unexpectedly, they did not find a correlation between breastmilk RNA and proviral DNA, suggesting that CFV may not derive from infected breastmilk cells but rather from local HIV replication in infected lymphocytes and macrophages or even by the migration of infected cells from the intestinal

mucosa to the mammary submucosa. Thus, compartmentalisation between breastmilk and peripheral blood showed that circulating viral populations alone may not be predictive of HIV postnatal transmission (Becquart *et al.*, 2002).

Conversely, other studies found limited to no evidence of compartmentalisation as demonstrated by extensive mixing of HIV-1 variants between breastmilk and blood (Gray *et al.*, 2011). Indeed, Gantt *et al.* and Salazar-Gonzalez *et al.* observed a high frequency of genetically identical and nearly identical clonally amplified variants in breastmilk (with higher clonal frequency as compared to blood), suggesting that relatively few infected cells in the mammary tissue transiently contribute a disproportionately high number of virions to the breastmilk viral pool (Salazar-Gonzalez *et al.*, 2011). HIV-infected cells that migrate to or reside within the mammary gland may contribute to the production of certain breast milk viruses, as indicated by newly identified clusters of clonally amplified variants closely related to concurrent blood virus variants. A lack of compartmentalization coupled with extensive clonal amplification of evolving functional viruses in breast milk, suggests continuous seeding of the mammary gland by blood virus variants, followed by transient local replication of these variants in the breastmilk compartment (Salazar-Gonzalez *et al.*, 2011). This is concordant with observations from Gantt *et al.* who suggested that compartmentalization based on monotypic or nearly identical sequences may not indicate a separate viral population but may rather reflect a snapshot of recent events such as a burst of viral replication and/or proliferation of HIV-1-infected cells (Gantt *et al.*, 2010).

Additionally, Gray *et al.* observed persistent multiple lineages of a strong population structure in plasma and breastmilk viruses (Gray *et al.*, 2011). They found that the time of the most recent common ancestor was before the start of lactogenesis, which is unexpected since breastmilk only becomes an active site of viral replication during lactogenesis. Thus, HIV may have migrated from another tissue, possibly from gut-associated lymphoid tissue (GALT) derived macrophages or T cells infected with an early HIV variant that migrated to the breast during lactogenesis and established infection. Cessation of ART may therefore result in the emergence of an early variant leading to long-lived cellular viral sources (Gray *et al.*, 2011). The discordance around viral variant compartmentalization could be related to factors such as differences in sampling intervals and viral load (Heath *et al.*, 2010).

1.4.3 Mastitis

Several trials have identified subclinical mastitis (SCM) as a significant risk factor for HIV-1 MTCT by breastfeeding (John *et al.*, 2001; Kantarci *et al.*, 2007; Lunney *et al.*, 2010; Rutagwera *et al.*, 2022; Semba *et al.*, 1999; Semrau *et al.*, 2008). For instance, Semba *et al.*, showed that women with increased breastmilk sodium levels exhibited twice the rate of HIV-1 MTCT as compared to those with normal sodium levels. They also showed that plasma and breastmilk viral load were linked to an increased risk of MTCT (Semba *et al.*, 1999). Interestingly, Semrau *et al.,* found that high breastmilk sodium concentration at 1 month postpartum was normal and unpredictive of HIV-1 MTCT as compared to four months postpartum which linked with increased viral load (Semrau *et al.*, 2008).

Other contributors of HIV-1 transmission by breastfeeding are CAV and CFV viral shedding which correlate positively with subclinical mastitis (Kantarci *et al.*, 2007; Rutagwera *et al.*, 2022). Rutagwera *et al*. reported a 7-fold increase of breastmilk HIV-1 CFV shedding from mild SCM to severe SCM with a median of 41 copies/ml and 309 copies/ml respectively, while Kantarci *et al.* observed a significant association between elevated breastmilk Na+/K+ (indicating SCM) and breastmilk HIV-1 CAV levels (Kantarci *et al.*, 2007; Rutagwera *et al.*, 2022). Rutagwera *et al.,* also observed an upregulation of proinflammatory cytokines in the mammary tissue accompanied by severe SCM which associated with increased viral shedding in breastmilk (Rutagwera *et al.*, 2022) which in turn strongly correlated with increased risk of HIV transmission (Semrau *et al.*, 2008).

Another study carried out by Gantt *et al.,* observed a positive correlation with breastmilk sodium concentrations and the mean genetic distance between breastmilk and blood variants. Additionally, mastitis was linked to increased HIV-1 viral load in breastmilk, likely due to an influx of plasma virions and local viral production within the breast, concordant with increased HIV-1 replication in breasts with mastitis (Gantt *et al.*, 2010).

1.5 The significance of breastmilk viral load in mother to child transmission

As previously mentioned, increased risk of postpartum vertical transmission by breastmilk is associated with increased maternal viral load ingested by infants (Njom Nlend, 2022). Hence, maternal prophylactic strategies are aimed at supressing viral loads to undetectable levels. Transmission may occur at any timepoint during breastfeeding, however early milk (colostrum) was found to harbour a higher viral load than mature milk, thus increasing transmission risk during early postnatal period (Rousseau *et al.,* 2003). Several trials have investigated the relationship between breastmilk viral load and the risk of HIV vertical transmission (Flynn *et al.*, 2018; Flynn *et al.*, 2021; Kuhn *et al.*, 2013; Rousseau *et al.*, 2003).

A particularly relevant study carried out by Rousseau *et al.* investigated the correlation between changes in breastmilk viral load over time and MTCT via breastfeeding (Rousseau *et al.*, 2003). They observed that breastfeeding mothers who transmitted HIV-1 to their infants demonstrated higher breastmilk RNA levels and consistent breastmilk viral shedding in comparison to those who did not transmit. They also showed that every 10-fold increase in breastmilk viral load correlated with a 2-fold increased risk of MTCT (Rousseau *et al.*, 2003).

These findings are congruent with another study conducted on the postpartum component of the International Maternal Peadiatric Adolescent AIDS Clinical Trials Group - Promoting Maternal Infant Survival Everywhere (IMPAACT PROMISE) trial which assessed the safety and efficacy of mART for the PMTCT in contrast to infant Nevirapine (iNVP) (Fowler *et al.*, 2016). In contrast to iNVP, mART significantly correlated with infant HIV-1 infection. Regarding women in the mART group, the incidence rate of infant infection was higher among women with baseline viral loads of more than or equal to 1000 copies/ml as compared to those below 1000 copies/ml. It was therefore concluded that among women under mART, high viral load during the breastfeeding period was linked to high risk of MTCT, highlighting the necessity of maternal viral suppression during breastfeeding, particularly in the absence of infant ART (Flynn *et al.*, 2018).

Consequently, maternal viral load monitoring during peripartum and postpartum period is essential as it is the best indicator of ART efficacy, adherence and enhanced infant postnatal ARV in case of high HIV acquisition risk (Ngandu *et al.*, 2022). This can be conducted by viral load testing during antenatal visits as well as retesting in the third trimester, peripartum and postpartum, as recommended by the WHO especially in regions experiencing high prevalence. In fact, seroconversion during late pregnancy is common whereby women are unaware of their new HIV status, often with unsuppressed viremia, thus increasing the risk of MTCT (Yang, Cambou and Nielsen-Saines, 2023).

As such, according to the latest updates (November 2019 guidelines), South Africa has set in place policies and testing schedules to monitor maternal viral load during different stages of the antepartum, peripartum and postpartum periods (Wessels *et al.*, 2020). Nonetheless, no known threshold of blood viral load has been identified below which transmission does not occur (Kourtis and Bulterys, 2010).

Therefore, breastmilk viral load testing may be a better predicator of the risk of HIV MTCT by breastfeeding as compared to plasma viral load testing. Despite plasma viral load reflecting breastmilk viral load in certain circumstances, HIV RNA may still be detectable in breastmilk and MTCT events by breastfeeding still occur despite suppressed viral load as measured in plasma (Goga *et al.*, 2021). However, routine monitoring of breastmilk viral load is not widely performed as it is costly and considered impractical (Goga *et al.*, 2021).

1.5.1 The burden of cell-associated virus (CAV) and cell-free virus (CFV) in breastmilk

CFV levels quantified by HIV-1 RNA and CAV measured by HIV-1 DNA have both been detected in breastmilk and correlate to the risk of MTCT (Koulinska *et al.*, 2006; Lehman *et al.*, 2008; Ndirangu *et al.*, 2012). Indeed, Koulinska *et al.* found that every 10-fold increase in CFV or CAV concentration correlated to about 3-fold increased risk of postpartum MTCT (Koulinska *et al.*, 2006).

Another observation was that CAV and CFV contribution to postpartum MTCT differed throughout the lactation period although it was unclear whether CAV or CFV was the main

contributor of vertical transmission by breastfeeding (Lehman *et al.*, 2008). Koulinska et al. quantified and compared breastmilk CFV and CAV levels in HIV transmitting versus non-transmitting mothers in Tanzania and reported detectable CFV levels in 57% of breastmilk samples and 75% of samples from transmitting mothers. They also detected CAV in 74% of all mothers as well as in 87% of mothers whose infants became infected postnatally. This study also demonstrated that breastmilk CAV load associated to MTCT before and after 9 months postpartum versus only after 9 months for breastmilk CFV (Koulinska *et al.*, 2006).

Similarly, a study conducted by Ndirangu *et al.* in HIV-infected mothers in KwaZulu-Natal observed that CAV correlated more strongly with transmission at 6 weeks postpartum while CFV was more strongly liked to MTCT at 6 months (Ndirangu *et al.*, 2012). Therefore, CAV may be used to predict the risk of MTCT during the early postnatal period while CFV may be an effective predictor in late postpartum transmission.

The aforementioned study also reported that triple ART regimen was effective at suppressing breastmilk CFV load, but CAV levels remained detectable in breastmilk (Ndirangu *et al.*, 2012). This was likewise highlighted by findings from Lehman *et al.,* who reported that CAV levels reduced by only 0.5 - 1 $\log_{10}$ per year of HAART treatment (Lehman *et al.*, 2008). Since breastmilk CAV is linked to HIV MTCT via breastfeeding, the persistence of CAV in breastmilk of both treated and untreated HIV-infected breastfeeding mothers may be a primary cause of HIV-1 transmission by breastfeeding (Koulinska *et al.*, 2006).

1.6 Antiretroviral therapy

1.6.1 Efficacy and limitations of ART for preventing HIV vertical transmission by breastfeeding

ART has demonstrated significant reduction of HIV transmission intrapartum, peripartum and postpartum (Bispo *et al.*, 2017; Kintu *et al.*, 2020; Lallemant *et al.*, 2020). ARV transferred through the placenta during gestation and via breastmilk during the postpartum period reduces the rate of HIV MTCT but transmission via breastfeeding persists despite maternal

prophylaxis (Davis *et al.*, 2019; Kesho Bora Study Group, 2011; Rousseau *et al.,* 2003). Indeed, a MTCT rate of 4.3% was recorded by 18 months of age where breastfeeding contributed to more than 80% of these transmissions during the first six months of the breastfeeding period (Wessels *et al.*, 2020). Despite undetectable HIV-1 viral load in blood, cell-free RNA may be present in breastmilk, as observed in the Breastfeeding, Antiretrovirals and Nutrition (BAN) study (Davis *et al.*, 2019). To date, no trial has been able to completely eliminate HIV from breastmilk despite the efforts to do so (Yang, Cambou and Nielsen-Saines, 2023).

For instance, the Kisumu Breastfeeding Study (KiBS), a single-arm trial carried out in Kenya investigated the safety and efficacy of triple mART to suppress viral load during the peripartum period (Thomas *et al.*, 2011). Under this study, HIV-positive women took Zdovudine, Lamivudine, and either Nevirapine or Nelfinavir from 34 - 36 weeks gestation to 6 months postpartum. Single-dose Nevirapine was administered to infants at birth and the women were advised to practice exclusive breastfeeding for 5.5 months with complete cessation of breastfeeding by 6 months (Thomas *et al.*, 2011). They observed an overall transmission rate of 3.2% from 6 weeks to 24 months postpartum, most probably due to breastfeeding. Nine infants tested HIV-positive after 6 months postpartum (in the absence of mART) of which only two mothers self-reported continued breastfeeding. Thus, unreported continued breastfeeding may have been a possible mechanism of late infection (Thomas *et al.*, 2011).

Another noteworthy study was the IMPAACT PROMISE trial (Flynn *et al.*, 2018). The PROMISE trial had previously demonstrated that mART during gestation and the intrapartum period could lower vertical transmission to about 0.5% in sub-Saharan Africa (Fowler *et al.*, 2016). Nonetheless, MTCT prevailed via breastfeeding in these areas and therefore, the PROMISE Postpartum Component was designed to compare the efficacy of extended infant ART versus mART for PMTCT by breastfeeding (Flynn *et al.*, 2018). HIV-positive women and their breastfeeding uninfected infants were randomized to either mART (Tenofovir, Emtricitabine and Ritonavir) or iNVP until 18 months postpartum or breastfeeding cessation, infant HIV-1 infection, or toxicity - whichever occurred first. They reported that overall postpartum MTCT was lower than expected in both arms, demonstrating the efficacy of both regimens with

postnatal infection rates of 0.6% and 0.9% at 12 and 24 months as well as over 97% of HIV-1-free survival at 24 months (Flynn *et al.*, 2018).

These results were similar to that of the Kesho Bora and the Breastfeeding, Antiretrovirals and Nutrition (BAN) trials, two studies, of equal significance to the prevention of MTCT via breastfeeding (Chasela *et al.*, 2010; Kesho Bora Study Group, 2011). Indeed, the Kesho Bora trial, conducted in Kenya (Burkino Faso), and South Africa enrolled and randomised pregnant women at weeks 28 - 36 of gestation to assess the safety and efficacy of option B triple ART (a combination of Zidovudine, Lamivudine and Lopinavir) until discontinuation of breastfeeding or at most 6.5 months postpartum in comparison to Zidovudine with single-dose Nevirapine (Kesho Bora Study Group, 2011). The triple ART group and Zidovudine with Nevirapine group reported overall transmission rates of 3.3% and 5.0% respectively at 6 weeks and 5.4% and 9.5% respectively at 12 months. Concerning mothers who planned to breastfeed, the overall MTCT rate was significantly lower at 5.6% in the triple ART arm as compared to 10.7% in the Zidovudine and Nevirapine arm at 12 months of age. At the same time point, maternal triple ART reduced the risk of vertical transmission by 43% compared to the Zidovudine and Nevirapine regimen (Kesho Bora Study Group, 2011).

Multiple studies have indicated that mART alone is insufficient to completely prevent MTCT (Davis *et al.*, 2016; Mangold *et al.*, 2021; Rousseau *et al.,* 2003; Thomas *et al.*, 2011). Nevertheless, integrase inhibitors such as Dolutegravir (DTG) may limit HIV integration in breastmilk cells, thus, potentially reducing the risk of cell-associated HIV MTCT through breastfeeding.

1.6.2 Efficacy of Dolutegravir at reducing HIV mother to child transmission

Many women in low to middle resource settings present with untreated HIV during late pregnancy (Myer *et al.*, 2017). Suppressed maternal viral load at the time of delivery is essential for the PMTCT however, conventional Efavirenz (EFV) based therapy does not suppress HIV viral load fast enough (Waitt *et al.*, 2019). Contrastingly, DTG has demonstrated rapid viral suppression in non-pregnant participants whereby the median time to reach viral suppression was 28 days with the DTG based ART group as compared to 84 days in the EFV-

based ART group (Walmsley *et al.*, 2013).

The IMPAACT 2010/VESTED open-label randomized trial compared the safety and efficacy of DTG compared to EFV based ART in pregnant women with confirmed HIV-1 infection from 14 to 28 weeks of gestation across North and South America, Asia and Africa (Lockman *et al.*, 2021). The trial evaluated peripartum safety and virologic efficacy of three different triple ART regimens administered during pregnancy, namely (i) a combination of DTG, Emtricitabine and Tenofovir Alafenamide Fumarate, (ii) a combination of DTG, Emtricitabine and Tenofovir Disoproxil Fumarate, and finally (iii) a combination of EFV, Emtricitabine and Tenofovir Disoproxil Fumarate. The trial showed that DTG based regimen achieved superior virologic suppression by delivery in contrast to EFV based ART (Lockman *et al.*, 2021).

Another important study that assessed DTG drug concentrations (including in breastmilk) and its effect on viral suppression was Dolutegravir in Pregnant mothers with HIV infection and their Neonates (DolPHIN-1). This open-label randomized clinical trial compared the use of a DTG to an EFV based standard of care regimen in HIV positive but treatment naïve women from Uganda and South Africa initiating ART in their third trimester of pregnancy (Waitt *et al.*, 2019). Pharmacokinetic analysis revealed high transplacental and modest trans-mammary transfer of DTG to infants. Coupled with delayed metabolic clearance of DTG in immature guts, infants experienced prolonged DTG exposure expected to last 6.3 - 7.9 days, based on a modelling analysis (Dickinson *et al.*, 2021; Waitt *et al.*, 2019).

Subsequently, a follow-up study, DolPHIN-2 reported that the primary efficacy outcome of viral load less than 50 copies/ml at the first postpartum visit was met by 74.2% of women on DTG and 42.7% of women on EFV based regimens and the median time to reach this target was 28 days and 82 days, respectively. Additionally, achieving viral load below 1000 copies/ml (secondary endpoint/efficacy outcome) took 7 days in the DTG group as compared to 23 days in the EFV group (**Figure 1.2**). Therefore, they concluded that DTG achieved viral suppression twice as fast as EFV based regimen when initiating ART in late pregnancy (Kintu *et al.*, 2020).

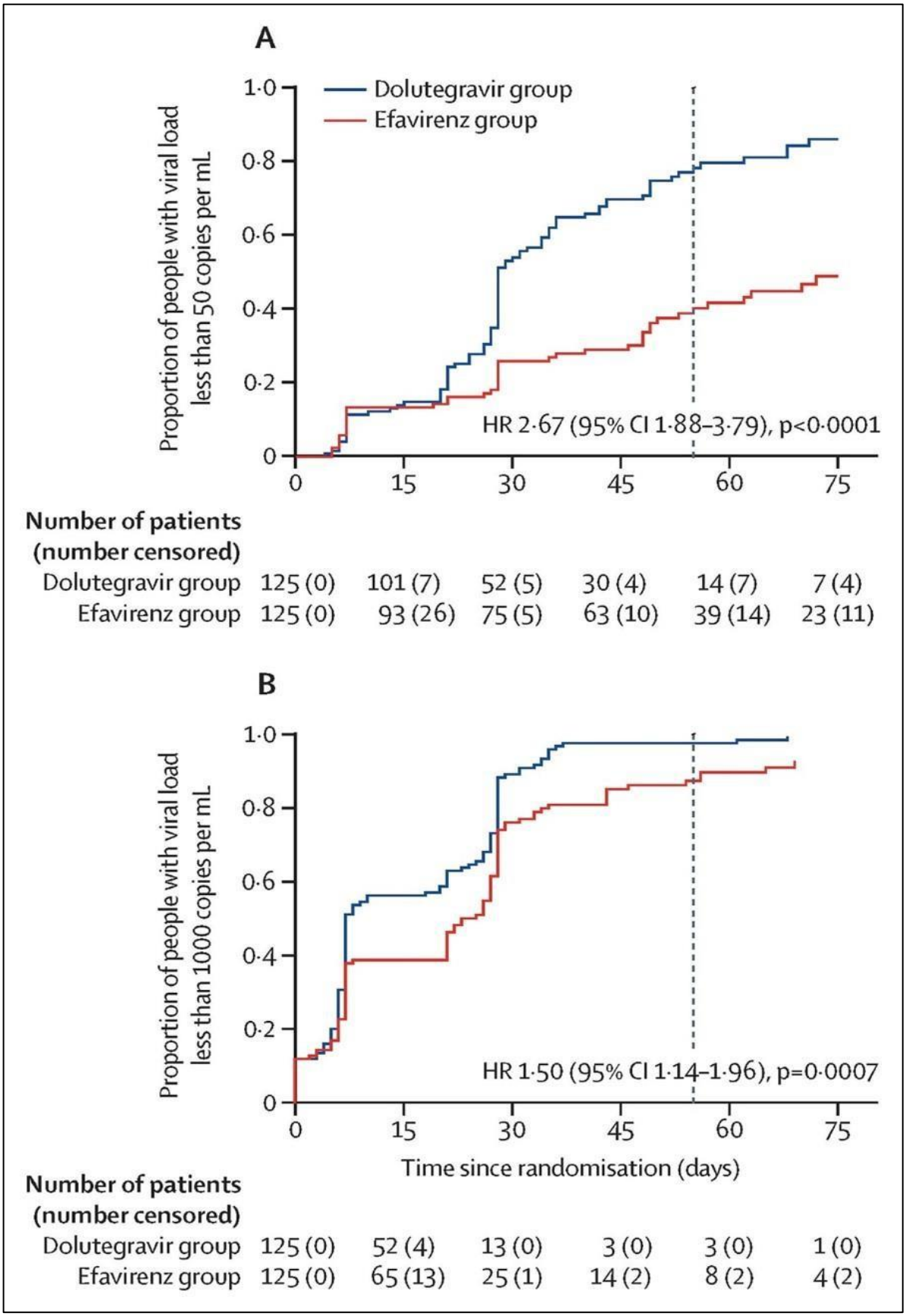
A
Dolutegravir group
Efavirenz group
Proportion of people with viral load less than 50 copies per mL
1·0
0·8
0·6
0·4
0·2
0
HR 2·67 (95% CI 1·88–3·79), p<0·0001
0
15
30
45
60
75
Number of patients (number censored)
Dolutegravir group 125 (0) 101 (7) 52 (5) 30 (4) 14 (7) 7 (4)
Efavirenz group 125 (0) 93 (26) 75 (5) 63 (10) 39 (14) 23 (11)
B
Proportion of people with viral load less than 1000 copies per mL
1·0
0·8
0·6
0·4
0·2
0
HR 1·50 (95% CI 1·14–1·96), p=0·0007
0
15
30
45
60
75
Time since randomisation (days)
Number of patients (number censored)
Dolutegravir group 125 (0) 52 (4) 13 (0) 3 (0) 3 (0) 1 (0)
Efavirenz group 125 (0) 65 (13) 25 (1) 14 (2) 8 (2) 4 (2)

Figure 1.2. Primary and secondary efficacy outcomes of the DolPHIN-2 study samples. Kaplan-Meier graphs illustrating time from randomisation to an HIV viral load of **A**. less than 50 copies/ml and **B**. less than 1000 copies/ml. The dotted vertical lines represent the median time since randomisation to childbirth (55 days) in all participants (Kintu *et al.*, 2020).

Regarding MTCT, three MTCT events were reported from the DTG arm, however, with low maternal viral loads at delivery and early infant positivity tests, thus in-utero transmission was the most probable path of MTCT, although peripartum transmission could not be excluded since the infants were not tested within 2 days of birth (Kintu *et al.*, 2020).

Additionally, one MTCT event occurred from the EFV arm between infant age of 12 - 72 weeks in spite of sustained maternal viral suppression and consecutive HIV negative infant tests during the first year of life. Transmission most probably occurred via breastmilk given the late positive test and that breastfeeding continued until 12 weeks postpartum. These results are in keeping with the presence of CAV reservoirs in breastmilk that persist despite ART or could be due to differences in plasma and breastmilk viral loads as well as HIV dynamics (Kintu *et al.*, 2020; Malaba *et al.*, 2022).

Overall, DTG was declared safe and effective during pregnancy as it rapidly reduces viral load to undetectable levels by the time of delivery and potentially during the breastfeeding period as well, thus potentially able to reduce HIV MTCT rates (Kintu *et al.*, 2020; WHO, 2019). Therefore, according to the current WHO guidelines for ART during pregnancy, a DTG based regimen combined with a nucleoside reverse-transcriptase inhibitor (NRTI) backbone is recommended as the first line ART (WHO, 2018). Nonetheless, the presence of CAV in women on DTG from the DolPHIN studies has not been investigated.

1.7 Breastmilk cells as viral reservoirs

1.7.1 Resting CD4+ T cells

Despite being very rare (about 10^3-10^7 latently infected CD4+ T cells per infected person) (Perelson *et al.*, 1996) latently infected CD4+ T cells remain unaffected by conventional ART (Finzi *et al.*, 1999). In breastmilk, a high ratio of HIV-1-infected cells to total breastmilk cells (BMC) is linked to HIV-1 transmission in the absence of ART, therefore if present in sufficient numbers, latently infected CD4+ T cells may contribute to HIV MTCT by breastfeeding during ART (Petitjean *et al.*, 2007). During the very early postpartum period, one millilitre (ml) of breastmilk was found to contain 3 million leukocytes (Crago *et al.*, 1979). Colostrum consists mainly of macrophages with 40-50% of total leukocytes, followed by neutrophils with 40-50% of total leukocytes and lymphocytes with 5-10% of total leukocytes of which about 83% are T cells (Hassiotou *et al.*, 2013). During the postpartum period, leukocyte levels decrease to 0–1.7% of total breastmilk cells in transitional milk and to 0–1.5% of total breastmilk cells in mature milk (Hassiotou *et al.*, 2013).

HIV-1 replication in the reservoir occurs when latently infected cells revert to productively infected cells upon spontaneous activation within the mammary gland possibly due to exposure to cytokines and chemokines present in the breastmilk microenvironment or during transepithelial migration/extravasation (Tuaillon, Valea *et al.*, 2009; Valea *et al.*, 2011; Wirt *et al.*, 1992). When activated, these cells can transcribe HIV DNA and generate virions in the mammary gland or in the infant's digestive tract (Van De Perre *et al.*, 2012). With an average of 2000 CD4+ T cells per ml of breastmilk, and the infant gut and tonsil mucosa exposed to an average of 700 ml of breastmilk every day, above a million maternal breastmilk CD4+ T cells can be ingested daily via breastfeeding. A child would have ingested $2x10^8$ CD4+T cells by breastfeeding at 6 months postpartum (Van De Perre *et al.*, 2012).

Activated memory T cells make up a major portion of breastmilk CD4+ T cells and the latter express higher levels of activation markers in comparison to their blood counterparts (Valea *et al.*, 2011). Indeed, latently infected cells in breastmilk may give rise to 10 times more HIV-1 antigen secreting cells (AgSCs) as compared to blood upon polyclonal activation (Becquart *et al.*, 2006; Van De Perre *et al.*, 2012). A high number of activated memory T cells are in fact

consistent with the physiological role of breastmilk as a source of immunologically active cells (Becquart *et al.*, 2006; Tuaillon, Al Tabaa, *et al.*, 2009). Valea *et al.* detected CD4+ activated memory T cells in blood and breastmilk despite undetectable HIV-1 RNA levels in these compartments during mART (Valea *et al.*, 2011).

Furthermore, a study conducted by Petitjean *et al.,* investigated the ability of latent HIV-1 infected resting CD4+ T lymphocytes in breastmilk to enter the replicative cycle (Petitjean *et al.*, 2007). Among the purified breastmilk cells, they analysed epithelial cell, the major cell population, and resting CD4+ T cell types. Whilst most epithelial cells were dead, contrastingly, resting CD4+ T cells represented above 90% of total purified and viable breastmilk cells. They also observed a larger T cell reservoir in breastmilk as compared to that in blood (median 400 and 57.14 HIV-1-AgSCs/10^6 resting CD4+ T cells, respectively). These observations illustrate a larger HIV-1 reservoir in breastmilk as compared to blood (Petitjean *et al.*, 2007). Therefore, breastmilk derived CD4+ T cells constitute a replication-competent form of the HIV-1 cellular reservoir and may be considered a viral sanctuary from which HIV may be released following polyclonal activation (Valea *et al.*, 2011; Van De Perre *et al.*, 2012).

1.7.2 Mammary Epithelial Cells

Mammary epithelial cells (MEC) have a unique functional role within the breast tissue. They are the milk-secreting cells of the mammary gland and in addition to being a protective layer, they select components from plasma to produce the nutrient and immune rich fluid that is breastmilk (Lawrence and Lawrence, 1999). Lymphocytes and monocytes/macrophages move paracellularly or transcellularly across a typically impermeable MEC barrier to reach the milk duct. Transcytosis of lymphocytes across the MEC may be caused by direct migration of lymphocytes through the MEC during pregnancy (Koulinska *et al.*, 2006). Since lymphocytes and MEC share physical contact, this may allow viral transmission from MEC harbouring HIV-1 to susceptible target cells in transit to milk such as leukocytes (Dorosko and Connor, 2010). Indeed, Dorosko and Connor found that MEC harboured HIV-1 virions in endosomal compartments and expressed HIV-1 receptor proteins CD4+, CCR5, CXCR4 despite a lack of evidence of direct MEC infection (Dorosko and Connor, 2010; Toniolo *et al.*, 1995). MEC were found to actively uptake and release infectious virions which facilitates infection and

replication in CD4+ target cells. Without the need for direct HIV-1 infection, MEC targeted cells that traffic across the mammary epithelial layer to breastmilk, hence contribute to viral transmission (Dorosko and Connor, 2010). Additionally, MEC seemed to form tight junctions with lymphocytes and even engulf intact lymphocytes in some cases, indicating close cell-to-cell communication between lymphocytes and MEC (Dorosko and Connor, 2010).

Viral transfer to target cells without *de novo* infection is further demonstrated by dendritic cell-specific intercellular adhesion molecule-3-grabbing non-integrin (DC-SIGN), a dendritic cell receptor that pathogens, including HIV-1 binds to, frequently expressed on breastmilk macrophages. The latter bind and capture HIV as well as assist its transport into breastmilk where infectious virions get transmitted to T cells by exocytosis, hence facilitating mucosal transmission (Satomi *et al.*, 2005) which increases the risk of HIV-1 MTCT by breastfeeding.

Moreover, MEC release secretions from its apical surface including a broad array of cytokines, chemokines, and growth factors, which enhance HIV-1 replication and proliferation of infected target cells thus further supporting these as an HIV viral reservoir (Dorosko and Connor, 2010).

Overall, since MEC actively participate in the uptake of HIV-1, facilitate CD4+ T cell infection, viral replication and proliferation and increase viability of infected quiescent T cells, it may act as a HIV-1 viral reservoir (Dorosko and Connor, 2010).

1.8 Quantifying HIV-1 DNA

1.8.1 Total HIV DNA

Defined as all infected cells and tissues containing all forms of HIV persistence, the HIV reservoir quantified by total HIV DNA biomarker may provide the most insight about pathogenesis (Avettand-Fènoël *et al.*, 2016). Quantification of total HIV DNA is simple, sensitive, standardized and reproducible due to polymerase chain reaction (PCR) platforms such as quantitative PCR (qPCR) or droplet digital PCR (ddPCR), which exhibit a large dynamic range and can accurately analyse large numbers of samples (Avettand-Fènoël *et al.*, 2016; Belmonti, Di Giambenedetto and Lombardi, 2021).

Total HIV DNA is thus the most widely used marker to explore the HIV reservoir, allowing the estimation of all forms of HIV DNA in infected cells (resting or activated) (Belmonti, Di Giambenedetto and Lombardi, 2021). Yet, total HIV DNA is considered an imperfect reservoir marker as it does not distinguish between integrated and non-integrated HIV genomes nor between defective forms from those that can produce infectious viruses. Therefore, PCR based assays that quantify total HIV DNA tend to overestimate the true size of the latent reservoir by amplifying defective or deleted proviruses which are not capable of producing virus in addition to integrated proviruses (Anderson and Maldarelli, 2018).

Nonetheless, total HIV DNA is clinically relevant as it influences the course of infection and can predict progression to Acquired Immunodeficiency Syndrome (AIDS) and death, irrespective of CD4+ T cell count and HIV RNA load. It can also predict cART response. Indeed, baseline HIV DNA remains quantifiable despite a decline during cART and reflects infection history as well as treatment efficacy (Avettand-Fènoël *et al.*, 2016). Total HIV-DNA load was also found to be the only parameter predictive of residual viremia and viral rebound (d'Ettorre *et al.*, 2010).

Therefore, despite its limitations, total HIV DNA quantification is generally accepted as a marker to measure the viral reservoir and monitor changes following specific treatment strategies.

1.8.2 Quantifying HIV-1 DNA in breastmilk

Multiple studies have explored HIV-1 DNA levels in breastmilk to inform transmission risk and efficacy of ART (Ghosh *et al.*, 2003; Kuhn *et al.*, 2013; Ndirangu *et al.*, 2012; Shapiro *et al.*, 2005; Slyker *et al.*, 2012) using qPCR methods. For instance, Shapiro *et al*. compared the breast milk viral loads of HIV-positive women on ART to that of untreated women. They found that HIV-1 DNA loads below 10 copies/10^6 cells was achieved in 50% of women on treatment. However, this low DNA concentration was also observed in 65% of untreated women (Shapiro *et al.*, 2005).

Ghosh *et al.* conducted a sensitivity analysis, similar to the first objective in our study (Ghosh *et al.*, 2003). They spiked cell-free milk with $\log_{10}$ serial dilutions of 8E5/LAV cells (each cell containing one proviral HIV DNA copy) while keeping the total number of cells in each sample

constant at 10^6 by adding peripheral blood mononuclear cells (PBMCs) or HIV-uninfected breastmilk cells. They detected down to 1 HIV copy in 20% of assays and could detect 10 HIV DNA copies with 80% sensitivity. Their data also suggested that breastmilk components such as lipids, immunoglobulins, glycoproteins, lactoferrin, and enzymes do not significantly inhibit PCR amplification (Ghosh *et al.*, 2003).

1.8.3 Platforms for HIV-1 DNA quantification

Throughout the years, several methods and assays have been developed to quantify HIV DNA in PLWH (Anderson and Maldarelli, 2018). Among these methods is real-time qPCR which is considered a gold standard for HIV detection. It is an indirect quantification method that depends on the quality of standard curves which is susceptible to minor changes in reaction efficiency (Trypsteen *et al.*, 2016). Instability in the curve may lead to quantification bias due to the logarithmic nature of PCR (Trypsteen *et al.*, 2016). Moreover, quantification cycle (Cq) values obtained from the standard curve are affected by mismatches between primer/probes and the individual's viral sequences as well as the presence of inhibitors and amount of DNA loaded (Rutsaert *et al.*, 2018). It was also observed that qPCR enhanced dynamic range at the cost of linearity and accuracy and exhibited high assay noise (Strain *et al.*, 2013).

Consequently, an improved, acute and standardized ddPCR-based method was optimized by Strain *et al.* to investigate the latent HIV reservoir in PBMCs or isolated CD4+ T cells (Strain *et al.*, 2013; Trypsteen *et al.*, 2016). ddPCR is based on the concept of limiting dilution where the DNA target, fluorescently labelled probe and the components of a PCR in a single bulk PCR reaction are partitioned into 20 000 nanoliter droplets and each droplet contains only one or no target DNA copy (Henrich *et al.*, 2012). Random distribution of the sample among droplets allows 20 000 PCR reactions from a single sample to take place in parallel (Anderson and Maldarelli, 2018) as illustrated in **Figure 1.3**. Large scale amplification allows precise and sensitive quantification of single template molecules. In order to reduce noise, signal from low-copy targets is amplified by decreasing signal from high-copy templates (BIO-RAD, no date; Strain *et al.*, 2013).

In contrast to qPCR, ddPCR allows direct absolute quantification of target molecules without the need for a standard curve generated by serial dilutions of known standards but instead,

uses Poisson statistics, reducing susceptibility to user error, hence rendering ddPCR more reproducible and precise than standard qPCR. ddPCR has been applied to several bodily compartments with a primary focus on HIV-1 DNA quantification (total HIV DNA) (Anderson and Maldarelli, 2018).

Strain *et al.* observed similar results from the two methods where the main contrast was structure and magnitude of noise (Strain *et al.*, 2013). Additionally, ddPCR demonstrated increased precision for the detection of HIV templates under a frequency of 300 copies/million cells. ddPCR showed increased tolerance to small changes in PCR efficiency (Trypsteen *et al.*, 2016) and primer/probe mismatches, resulting in more accurate quantification as compared to qPCR (Trypsteen *et al.*, 2016) and rendering it an ideal platform for HIV quantification since sequence heterogeneity is common (Trypsteen *et al.*, 2016). ddPCR also showed reliable quantification of proviral and episomal HIV-1 DNA targets below qPCR's limit of quantification with an absolute limit of detection ranging from 0.7 to 3 copies per million CD4+ T cells (Strain *et al.*, 2013). Demonstrating increased sensitivity, single targets as small as 60-80 nucleotides have previously been detected by ddPCR, thus, rendering it an ideal platform for investigating rare events and its broad linear range of quantification allows resolution of small changes in abundance with sensitive two-fold difference detection (Anderson and Maldarelli, 2018). Furthermore, the ability of ddPCR to handle comparably large amounts of input DNA despite PCR inhibitors renders it ideal for the investigation of very low levels of viral genetic material such as HIV-1 reservoirs (Henrich *et al.*, 2012). While false positive events have been reported for this platform (Strain *et al.*, 2013), it is still likely superior to traditional qPCR due to the aforementioned reasons. The ddPCR workflow is illustrated below (**Figure 1.3**).

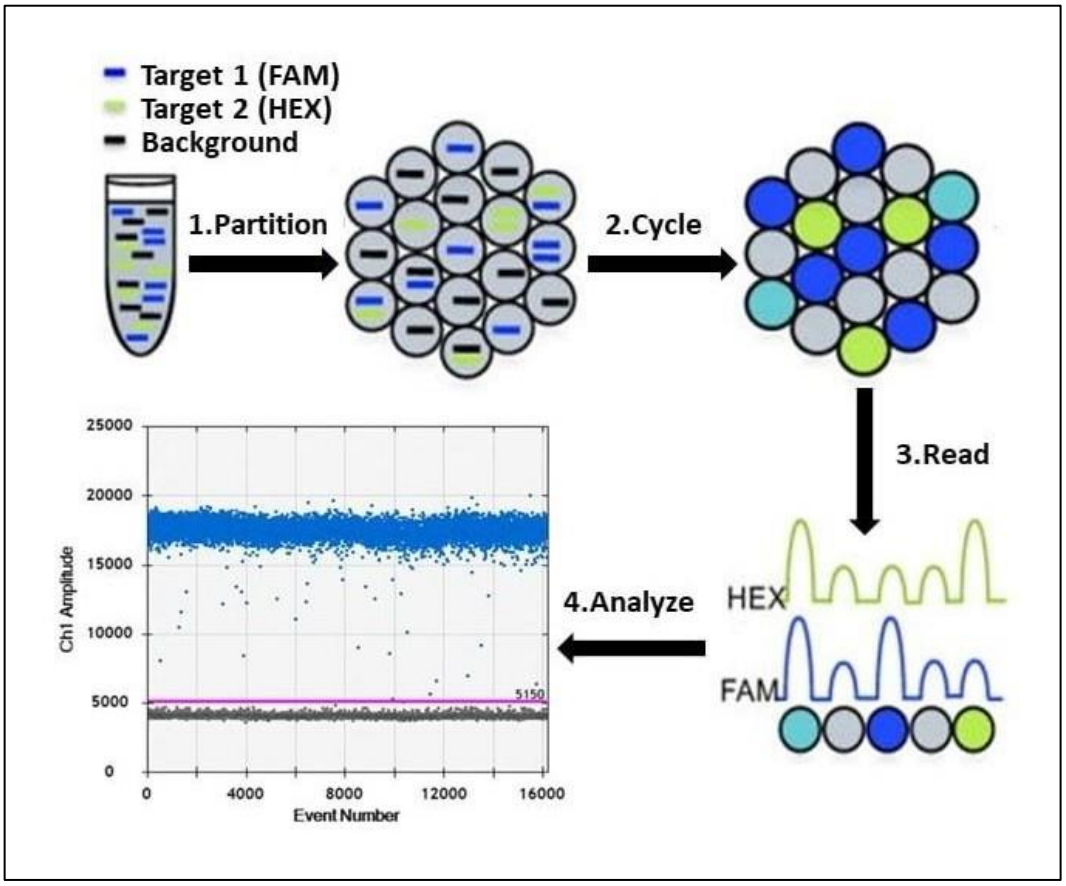

Figure 1.3. ddPCR workflow illustrating PCR mastermix containing DNA (target and background) along with primers and fluorescently labelled probes. With the help of droplet oil, the mixture gets partitioned into 20 000 nanoliter sized droplets with the QX200 droplet generator followed by target amplification in a C1000 Touch thermal cycler. A QX200 droplet reader autosampler picks up droplets from each well such that the droplets are streamed in a single file individually past a two-color fluorescence detector (e.g. FAM and HEX) which measures droplet-by-droplet fluorescence of each sample in two channels. The data generated is analysed in the QuantaSoft™ software where thresholds are set between the negative and positive droplets (Anderson and Maldarelli, 2018).

To our knowledge, HIV-1 DNA quantification in breastmilk samples using the ddPCR platform is a novel approach. This may be beneficial to evaluate HIV persistence in mothers on ART and its contribution to MTCT and more broadly, breastmilk as an HIV reservoir.

1.9 Sequencing HIV-1 DNA in breastmilk

Viral sequence information is also needed to support evidence for MTCT via breastfeeding from mothers on ART as well as characterisation of breastmilk as a reservoir.

1.9.1 Sanger sequencing

HIV infection is known for its extensive genetic diversity and rapid evolution across infected individuals (van Zyl, Bale and Kearney, 2018). Previous studies have sequenced breastmilk RNA and DNA to explore drug resistance, the transmission bottleneck, and compartmentalization. These investigations have mostly relied on the traditional Sanger sequencing method, which is limited in its ability to detect low-frequency variants and fully capture the breadth of viral diversity (Danaviah *et al.*, 2015; Heath *et al.*, 2010; Koulinska *et al.*, 2006; Nakamura *et al.*, 2017). Some of these sequence studies are described below.

One study carried out by Becquart *et al.* compared HIV-1 viral populations in different fractions of breastmilk and between each breast. They analysed the hypervariable env C2-V5 region from breastmilk samples of South African HIV-positive mothers unexposed to ART and found that viral DNA and RNA (reflecting virions) populations were distinct from each other. Using this approach they were able to propose that proviral or virion forms of HIV-1 in breastmilk are both susceptible to a specific microenvironment, thus driving quasispecies viral selection and that each mammary gland is subjected to microenvironmental pressure that may differ from the other breast (Becquart *et al.*, 2007).

Another study carried by Danaviah *et al.* explored the origin of HIV transmission during breastfeeding as well as the link with breastmilk CFV and CAV and maternal plasma variants with intrapartum single dose Nevirapine administered to mothers and infants (Danaviah *et al.*, 2015). While both CFV and CAV contributed to breastmilk transmission, the breast was mainly seeded by a single viral variant (founder virus). The most recent common ancestor of the major HIV variant was found to have established in the breast tissue during mammary development (months before lactation), during the current or a previous pregnancy, as formerly observed by Gray *et al.*, thus reflecting the presence and influence of long-lived

ancestral lineages within the breast (Danaviah *et al.*, 2015; Gray *et al.*, 2011). They found also that the genetic bottleneck that forms during HIV-1 transmission by breastfeeding seemed to be mediated by selection from within the breast in 50% of their cases (Danaviah *et al.*, 2015). However, comparison of HIV-1 populations between blood and breastmilk samples have shown contradictory evidence for compartmentalization in the mammary tissue as previously discussed (section 1.4.2).

1.9.2 Next Generation Sequencing

Several NGS platforms are currently available, such as the second-generation approach (e.g. Illumina and Ion torrent) that allows parallel sequencing of short reads while the third-generation approach (e.g. Pacific Biosciences and Oxford Nanopore technologies) can achieve read lengths exceeding 10 kb.

A widely used NGS platform, which is also at the centre of this research project, is the Illumina MiSeq. It is a rapid, high-throughput process that allows massively parallel sequencing of single DNA molecules resulting in a significant increase in data output as compared to Sanger sequencing (Hu *et al.*, 2021), which renders it a cost-effective platform for sequencing large target numbers. Indeed, NGS sequences hundreds to thousands of genes simultaneously to generate billions of reads while Sanger sequencing sequences only one DNA fragment at a time (Illumina, 2024). With Sanger sequencing having a high limit of detection (~15–20%) and low discovery power, NGS offers increased sequencing depth and enhances sensitivity to detect novel or rare variants at levels below 20% (Illumina, 2017; Ouyang *et al.*, 2024). Hence, Illumina Miseq may be an ideal platform to sequence low-level HIV variants from the viral reservoir as usually observed with PLWH who are on ART. The Illumina Miseq protocol is illustrated below (**Figure 1.4**).

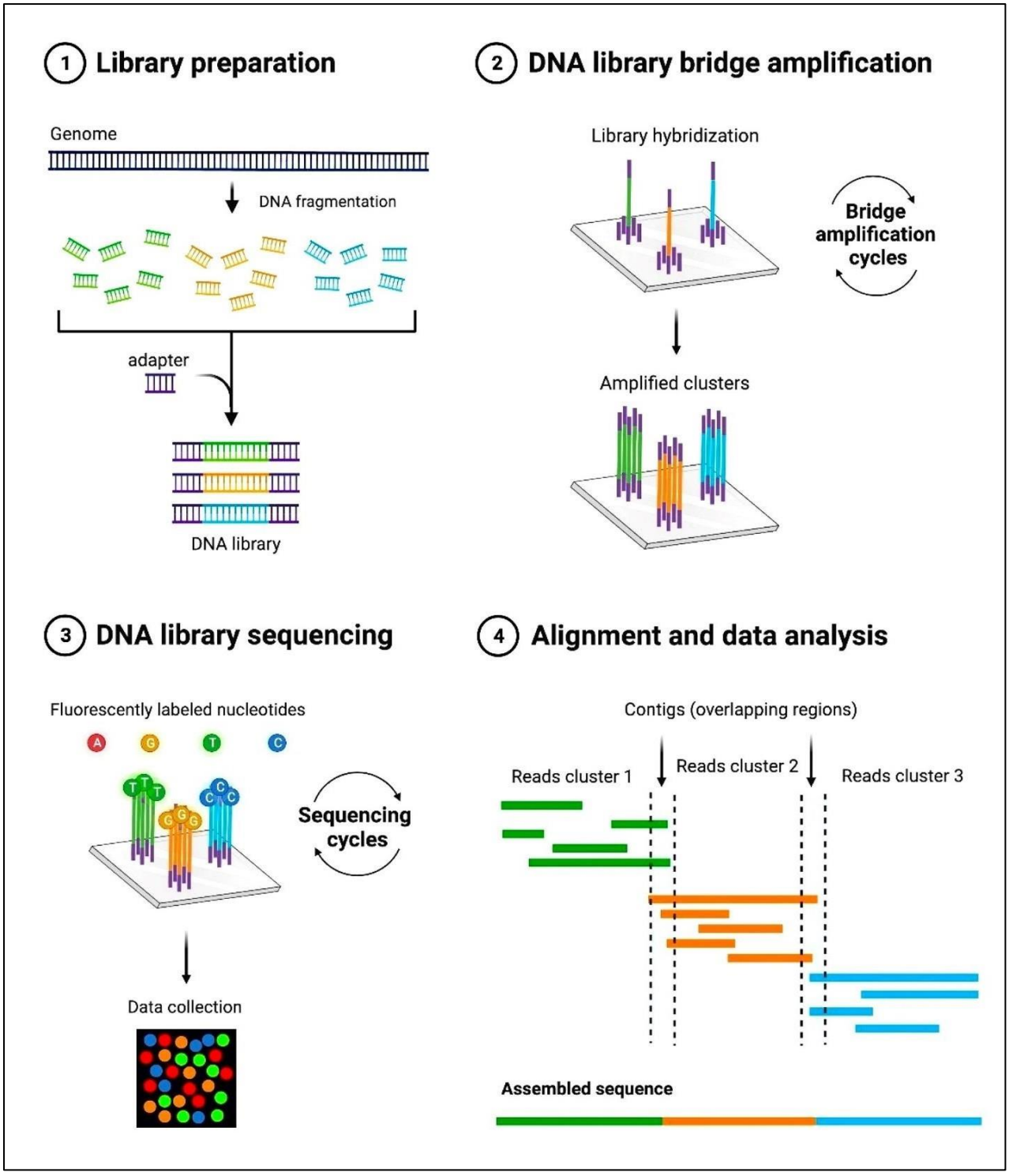

Figure 1.4. Illumina Miseq protocol. First, a genomic DNA (gDNA) or cDNA library is prepared. This involves DNA sample fragmentation and addition of adapters to each fragment end followed by denaturation to produce single strand DNA (ssDNA) fragments. Second, the ssDNA hybridise with oligonucleotides on the flow cell. ssDNA fold and hybridise again and form a double stranded bridge which gets denatured, resulting in two identical fragments

bound to the flow cell. This step is repeated to form millions of clusters. Third, sequencing by synthesis follows where each strand binds to one uniquely coloured fluorescently labelled nucleotide at a time. The sequence of distinct colours determines the nucleotide sequence. All identical strands are read simultaneously for each cluster. Fourth, data analysis is performed by overlaying sequences, pairing forward and reverse reads and alignment with a reference genome for variant identification (Esposito, Esposito and Ptashnik, 2022; Illumina, 2017).

NGS has been used in several HIV MTCT studies but mainly to investigate drug resistance (Acosta *et al.*, 2021; Fisher *et al.*, 2015; Samuel *et al.*, 2014). With regard to HIV MTCT by breastfeeding, we found no studies which applied NGS to sequence HIV DNA from breastmilk. Therefore, we are proposing to apply NGS to address this gap in research and provide valuable information on HIV MTCT by breastfeeding that has not yet been generated.

1.10 Rationale

Despite adherence to ART and successful viral suppression in mothers postpartum, instances of vertical transmission persist, primarily due to the presence of latently HIV-infected cells in breastmilk. Integrase inhibitor DTG has demonstrated promise at reducing the risk of transmission via breastfeeding by rapidly inhibiting HIV integration, however, its ability to reduce CAV load is yet to be quantified. Preventing integration of proviral DNA may halt the formation of a latent HIV reservoir. There is a pressing need for reproducible methods to assess HIV persistence within the breastmilk reservoir. Methods such as droplet digital PCR and next-generation sequencing may be applied to breastmilk from HIV-infected mothers to inform the risk of HIV transmission by breastmilk as well as the efficacy of drug trials.

1.11 Aim

To optimize a droplet digital PCR and next-generation sequencing method to respectively detect and sequence low-level HIV-DNA in breastmilk samples from breastfeeding women on antiretroviral treatment.

1.12 Objectives

1. Evaluate sensitivity of HIV DNA copy number detection using droplet digital PCR on HIV negative breastmilk samples spiked with cells harbouring the HIV genome.
2. Quantify HIV DNA copy number in breastmilk cells from women from the DolPHIN-2 study.
3. Evaluate sensitivity to sequence HIV from breastmilk cells using the Illumina MiSeq next generation sequencing platform.

Chapter 2: Materials and Methods

2.1 Ethics statement

This study was approved by the University of Cape Town's Faculty of Health Sciences Human Research Ethics Committee (UCT FHS HREC) (reference number: 559/2021). It is a sub-study of the DolPHIN-2 (Dolutegravir in pregnant mothers with HIV infection and their neonates 2) trial (HREC reference number: 096/2017).

2.2 Experimental Samples

2.2.1 HIV uninfected breastmilk samples

Breastmilk samples from HIV uninfected women were collected for spike-in with 8E5 cells, each containing a single copy of HIV-1 defective genome (spike-in described in section 2.4). Breastmilk samples from anonymous, HIV uninfected, breastfeeding women obtained from the Mowbray breastmilk bank in Cape Town were collected from Groote Schuur hospital following pasteurization and sterility testing. The Holder pasteurisation method was used whereby milk is heated to 62.5°C and held at that temperature for 30 minutes followed by rapid cooling to 4°C. The samples were thereafter frozen at -20°C and sent for sterility testing on the following day. Pasteurization and sterility testing were performed as routine processing of breastmilk samples at the Groote Schuur Hospital diagnostics unit. A total of seven breastmilk samples were collected. Neither demographic nor personal information was available for the donors. The milk samples were swirled to mix as lipid layers usually settle on top of the aqueous layer when left to stand. Then, 20 ml of each milk sample was decanted into labelled 50 ml conical tubes, which were transported on ice and stored at 4°C. To determine cell counts and viability, breastmilk samples were centrifuged at 14 000 rpm for 10 minutes followed by removal of the fatty layer and supernatant. The cell pellets were resuspended in 10 ml of Roswell Park Memorial Institute (RPMI-1640) media (Thermo Fisher Scientific, Waltham, USA) and their cell count and viability assessed using trypan blue dye and a Countess™ automated cell counter (Invitrogen, Carlsbad, USA).

2.2.2 DolPHIN-2 clinical trial breastmilk samples (for detection of low-level HIV-1 DNA)

Breastmilk cell pellets from 2 ml of breastmilk from each of six DolPHIN-2 study participants were obtained in duplicate from the DolPHIN-2 clinical trial (Kintu *et al.*, 2020). The cell pellets were provided by Mr. Nai-Chung Hu and stored at -80°C followed by resuspension in PBS prior to DNA extraction as described below (section 2.5).

2.2.3 CAPRISA study samples (for sequencing optimisation)

Genomic DNA from PBMCs of three CAPRISA 002 study participants (Van Loggerenberg *et al.*, 2008) on antiretroviral treatment and selected based on previously quantified proviral DNA copy number, were obtained from Dr. Sherazaan Ismail, Division of medical virology, Faculty of Health Sciences, UCT.

2.3 Cell culture

The 8E5 clonal cell line (ATCC CRL-8993), consists of lymphocytes containing a single, reverse transcriptase-defective copy of an integrated HIV-1 genome (Powell *et al.*, 1988). This prevents virion production, rendering it a safer alternative for HIV studies. The reagent was obtained through the NIH HIV Reagent Program, Division of AIDS, NIAID, NIH: Human Immunodeficiency Virus 1 (HIV-1) Lymphadenopathy-Associated Virus (LAV)-Infected 8E5 Cells, ARP-95, contributed by Dr. Thomas Folks.

Cryopreserved 8E5 cells and PBMCs were removed from liquid nitrogen storage, rapidly thawed in a 37°C water bath and washed in 30 ml of complete RPMI-1640 media (Thermo Fisher Scientific, Waltham, USA). Centrifugation at 21 000 rpm for 8 minutes was performed, followed by removal of the supernatant and resuspension of the cell pellet in 10 ml of complete RPMI-1640 media (supplemented with 10% Foetal Calf Serum, 0.5% gentamicin and 0.1% amphotericin B). Cell count and viability were measured using 0.2% Trypan blue stain (Thermo Fisher Scientific, Waltham, USA) on the Countess™ automated cell counter. PBMCs were resuspended at 1 million cells per ml and maintained in complete RPMI-1640 media at 37°C with 5% $C0_2$ for the period of the spike-in experiment. 8E5 cells were split at a

concentration of 100 000 cells/ml and cultured at 37˚C with 5% $C0_2$. 8E5 cells were passaged approximately every three days over a period of two months to maintain viability.

2.4 Spike-in of breastmilk samples.

We performed three distinct spike-in experiments using breastmilk samples, 8E5 cells and background cells. Breastmilk samples were spiked with different combinations of cells including 8E5 cells only, 8E5 cells and PBMCs as well as 8E5 cells and SUP-T1 cells.

2.4.1 Spike-in of breastmilk with 8E5 cells only

A 10-fold serial dilution of 8E5 cells in complete RPMI-1640 media was performed starting from an initial concentration of 100 8E5 cells/µl down to 0.01 8E5 cells/µl. 100 µl of each dilution was then spiked into aliquots of breastmilk for which cell count and viability were determined. In order to keep the total number of cells (breastmilk cells and 8E5) in each sample constant at one million, the required number of breastmilk cells was calculated and the corresponding volume of milk was subsequently used to spike into. In this way, an expected spike-in range of 1 to 10 000 8E5 cells per million cells total was achieved (**Figure 2.1**). Breastmilk samples BM1, BM2 and BM3 were used in this process and spike-ins were performed in duplicate for each sample (**Table A1**, Appendices). The spike-in workflow is depicted in **Figure 2.1.**

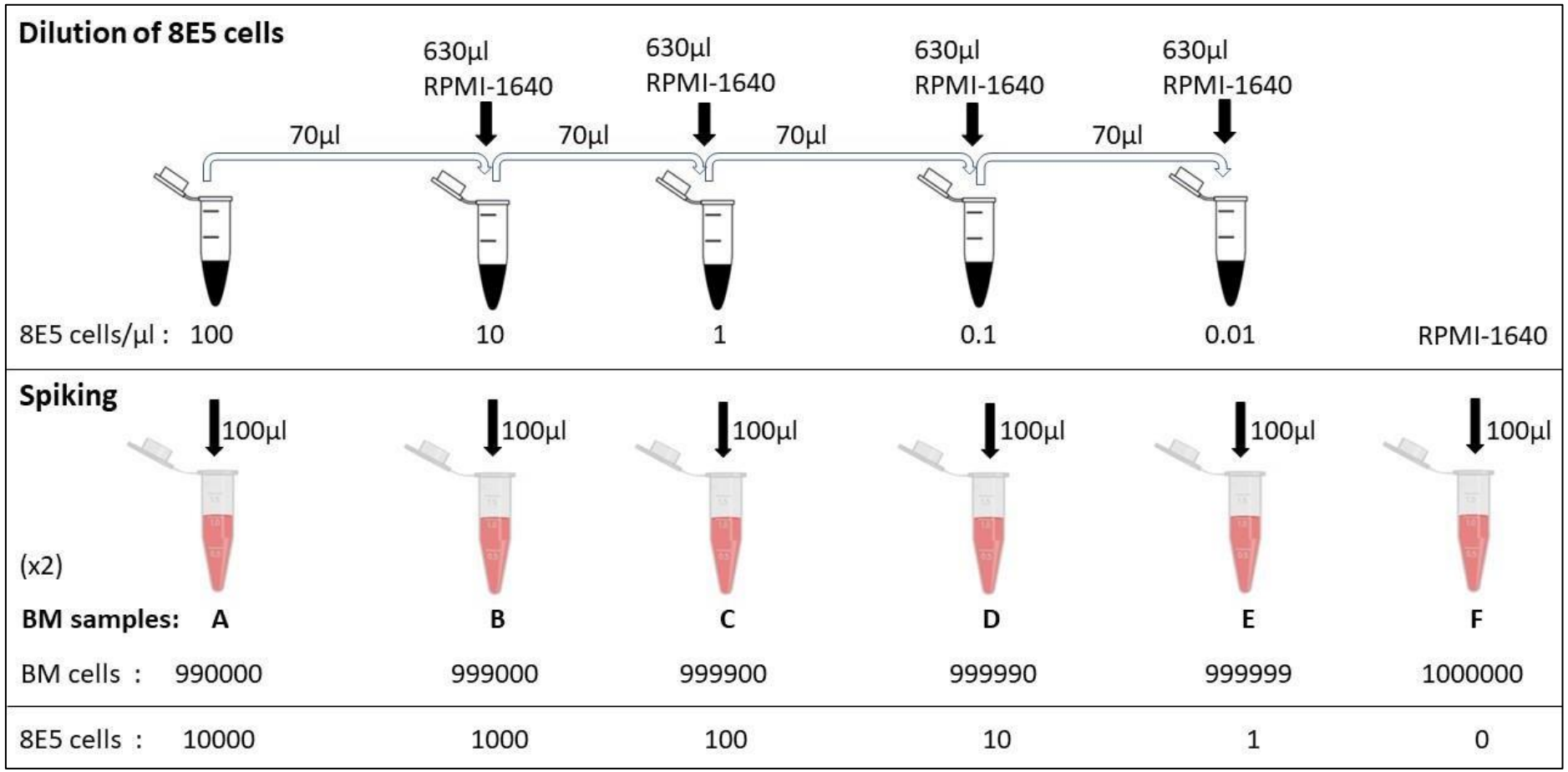

Figure 2.1. Spike-in of breastmilk samples BM1, BM2 and BM3 with decreasing concentrations of 8E5 cells ranging from 100 cells/µl to 0.01 cell/µl (A-E). Sample F was a non-spiked negative control to which 100 µl RPMI-1640 media was added. This was carried out in duplicate for each breastmilk sample.

2.4.2 Spike-in breastmilk with 8E5 cells and PBMCs

8E5 cells were diluted as previously described in section 2.4.1 and 100 µl of each dilution was spiked into 1 ml aliquots of breastmilk. The breastmilk cell count was not considered in this case. In order to keep the total number of cells (PBMCs and 8E5) in each sample constant at one million, the required numbers of PBMCs were calculated and the corresponding volume of resuspended PBMCs was subsequently spiked in (approx. 354µl of 2.8 x 10^7 PBMCs per 10 ml RPMI-1640 media). This procedure was performed with breastmilk samples BM4 and BM6. Spike-ins were carried out in duplicate for each breastmilk sample (**Table A2**, Appendices). The spike-in workflow is depicted in **Figure 2.2**.

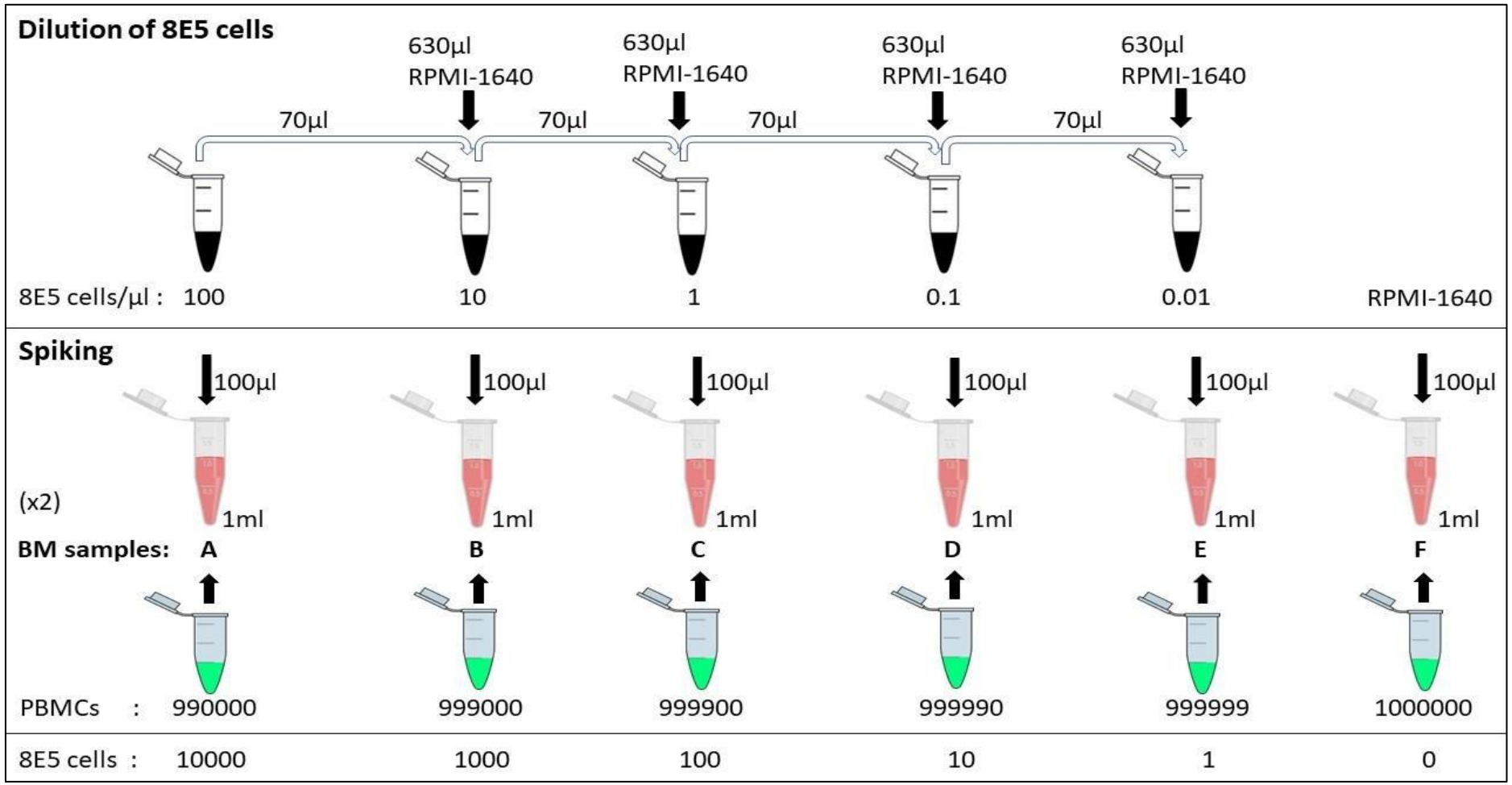

Figure 2.2. Spike-in of breastmilk samples BM4 and BM6 with decreasing concentrations of 8E5 cells ranging from 100 cells/µl to 0.01 cell/µl respectively (A-E) as well as PBMCs. Sample F was a non- 8E5 spiked negative control to which 100 µl RPMI-1640 media and PBMCs were added. This was carried out in duplicate for each breastmilk sample.

2.4.3 Spike-in breastmilk with 8E5 cells and SUP-T1 cells

Spike-in of HIV naïve breastmilk samples was repeated as described in the section 2.4.2 above with some modifications as shown in **Figure 2.3**. For instance, to account for variability across samples, 10 ml of each of three donor breastmilk samples were pooled together and mixed. Additionally, the 8E5 cell dilutions ranged from 500 cells/µl to 0.05 cell/µl, following a 10-fold serial dilution. The total number of cells (disregarding breastmilk cell count) was increased to 5 million, supplemented with SUP-T1 CD4 T-cells (CRL-1942) (American Type Culture Collection [ATCC], 2024 ; Smith *et al.*, 1988). SUP-T1 cells were obtained from the NIH HIV Reagent Program, BEI Resources (NIH/NIAID). The number of SUP-T1 cells required for each sample A-F was calculated, their corresponding volumes aliquoted in 15 ml tubes and pelleted by centrifuging at 21 000 rpm for 8 minutes. The supernatant was then removed and each pellet resuspended directly in 1 ml of pooled breastmilk sample BM8. Subsequently, 500 µl of each 8E5 dilution was spiked into 1 ml of SUP-T1 spiked breastmilk samples. Spike-ins were carried out in duplicate (see **Table A3**, Appendices).

A calculated proviral copy number concentration of 0.8 copies/8E5 cell was used. For instance, sample BM8A spiked with 250 000 8E5 cells was expected to contain 200 000 provirus copies, as shown in **Figure 2.3**.

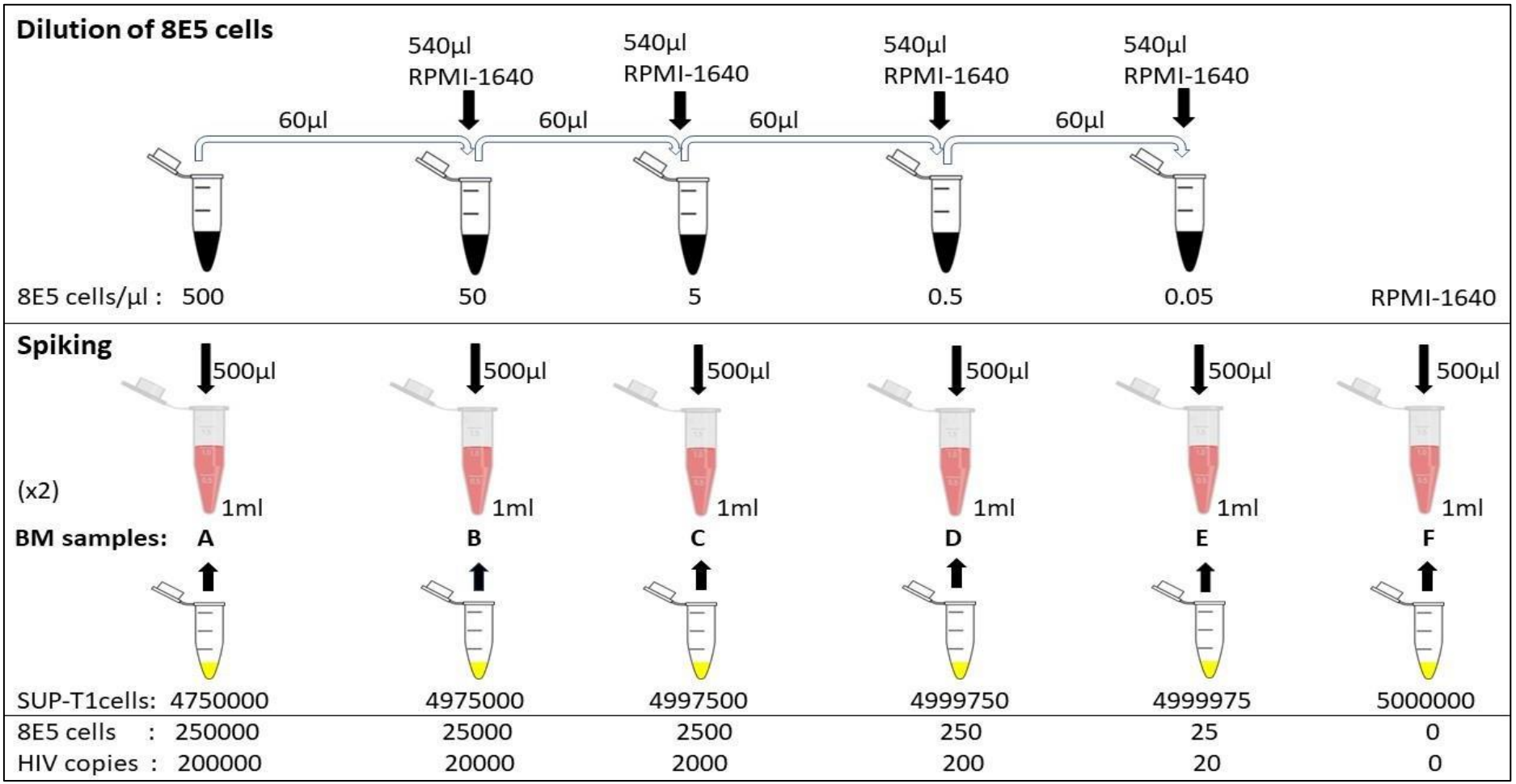

Figure 2.3. Spike-in of breastmilk samples BM8 with decreasing concentrations of 8E5 cells ranging from 500 cells/µl to 0.05 cell/µl respectively (A-E) as well as SUP-T1 cells. Sample F was a non- 8E5 spiked negative control to which 500 µl RPMI-1640 media and SUP-T1 cells were added. This was carried out in duplicate for each breastmilk sample.

2.5 DNA extraction

gDNA was extracted using the QIAGEN AllPrep DNA/RNA Mini Kit (QIAGEN, Hilden, Denmark), following the manufacturer's protocol. The extraction included the use of Proteinase K to digest contaminating proteins from the sample. Buffer AL, containing a guanidine salt, was added for lysis of cell membranes as well as inactivation of nucleases and promoting of DNA binding to the spin column. Washes with ethanol and buffers AW1 and AW2 were performed to purify the DNA. Elution with 200 µl EB buffer twice allowed the maximum yield of DNA from the column. This was followed by ethanol precipitation to increase DNA concentration. Four volumes of absolute ethanol (1600 µl) and one tenth volume (40 µl) of 3M sodium acetate at pH 5.5 were added to the eluate, followed by mixing by inversion and incubation at -20°C overnight. Post incubation, centrifugation was carried out at 14 000 rpm at 4°C for 30 minutes and the supernatant removed. 1 ml of 70% ethanol was then added and the samples were centrifuged again under the same conditions. Position of the pellet remained unchanged during spinning. The DNA pellets were air dried and resuspended in 30 µl molecular grade water and their concentrations measured using a NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific, Waltham, USA) at absorbance wavelengths of 230 nm, 260 nm and 280 nm.

2.6 Droplet digital PCR

2.6.1 ddPCR primers and probes compatibility analysis

AliView™ v 1.28 (Larsson, 2014) was used to generate multiple sequence alignments to identify compatibility between existing HIV-1 subtype C ddPCR primers and probes and 8E5 subtype B HIV target regions.

Subtype C-specific primers and probes were designed for 2-long-terminal repeat (2-LTR) circles using the Sequence Locator (Los Alamos National Laboratory [LANL], 2019) and OligoAnalyzer™ (Integrated DNA Technologies [IDT], 2024) tools based on previously reported subtype B counterparts (Strain *et al.*, 2013). Briefly, the 2-LTR primers and probes

described by Strain *et al*. were aligned and compared to an HIV subtype C consensus sequence using the AnalyzeAlign™ tool (LANL, 2020). The consensus subtype C sequences were in turn analysed via the OligoAnalyzer™ tool and checked for melting temperatures, GC content, delta G values and secondary structures.

HIV subtype B primers and probes as reported by Strain et al. were used to amplify conserved regions of *pol* and *gag* genes of 8E5 gDNA (Strain *et al*., 2013). *rpp30* primers and probes were also used to amplify the housekeeping gene Ribonuclease P protein subunit p30 (*rpp30)* for all samples. With two gene copies in each diploid cell (Dyavar *et al.*, 2018), *rpp30* is widely expressed in immune cells and several other tissues (The Human Protein Atlas, 2022). *rpp30* is highly conserved, rendering it a common control for copy number studies (Rowlands *et al.*, 2019). For ddPCR of subtype C samples, subtype C-specific *gag* and *pol* primers and probes were used (**Table A4**, Appendices).

2.6.2 ddPCR workflow

The HIV Diversity Group standard operating procedure (SOP) (HDL_P3.0SOP44.1) was followed. Each reaction mixture contained 10 µl of ddPCR Supermix for probes (no dUTP) (BIO-RAD, California, USA), 0.18 µl of 50 ng/µl *gag* and/or *pol* primers, 0.2 µl of 25 ng/µl *gag* and/or *pol* probe, 0.50 µl of Ban II restriction enzyme (New England Biolabs, Massachusetts, USA), 0.70 µl of CutSmart buffer (New England Biolabs, Massachusetts, USA), 20 ng – 1000 ng of template DNA, and molecular grade water up to a total volume of 20 µl (see sections 2.2 and 2.3 in Appendices). Incubation at 37˚C for 30 minutes was carried out to digest the gDNA. This allowed separation of tandem gene copies, reduced sample viscosity and improved template accessibility for input samples greater than 66 ng per reaction for overall optimised accuracy. Reaction mixtures along with droplet generator oil for probes (BIO-RAD) in cartridges were then loaded into the BIO-RAD QX200™ droplet generator which partitions each sample into up to 20 000 oil droplets. The droplets were then loaded onto a 96-well PCR plate and heat-sealed using a PX1 PCR plate sealer. This was followed by DNA amplification in the BIO-RAD C1000 touch thermal cycler or the Applied Biosystems™ thermal cycler 2720 (Thermo Fisher Scientific, Waltham, USA) under the following cycling conditions: 10 minutes

at 95°C, 50 cycles each consisting of a 30 second denaturation at 94°C followed by a 60°C annealing and extension for 60 seconds, and lastly, 10 minutes at 98°C. Furthermore, ramp rate was set to 2°C/s, volume to 40 µl and lid to 105°C. The droplets were analysed directly after cycling using the BIO-RAD QX100™ droplet reader or stored at 4°C overnight until analysis.

2.6.3 ddPCR analysis

ddPCR data were analysed using the QuantaSoft™ Analysis Pro v. 1.0 software (BIO-RAD, 2016) where amplitude thresholds were set conservatively per experiment using control wells to exclude false positive and background signals. Raw data were visualized as positive and negative droplet populations differentiated by amplitude of probe fluorescence (FAM or HEX). Template concentration was calculated per wells as target copies/µl. Wells with less than 9000 droplets or droplets with inconsistent amplitude signals were excluded from analysis. Wells with target concentrations below the Poisson confidence maximum of non-template and negative controls were assigned a concentration of 0 target copies/µl. Post normalisation to target copies per 1 µg DNA input for both *gag*/ *pol* and *rpp30*, reservoir size of target copies/ million cells was calculated as follows:

$$\frac{gag\ or\ pol\ target\ copies/\mu l}{\left(\frac{rpp30\ copies/\mu l}{2}\right)}\ x\ 10^6$$

Additionally, bar charts illustrating ddPCR target concentrations for each sample were generated using GraphPad Prism v5.00 (GraphPad Software, 2007).

2.7 Illumina MiSeq PCR

2.7.1 Pre-PCR restriction enzyme digest

Due to the large size and viscosity of gDNA, a restriction enzyme digest (see **Table A16** in Appendices) was performed prior to PCR, aimed to digest a maximum of 1000 ng of template DNA at varying sample concentrations. Therefore, each reaction contained 2 µl of CutSmart buffer at 10x concentration as well as 0.5 µl of Ban II restriction enzyme (New England Biolabs, Massachusetts, USA), 0 - 17.5 µl of template DNA and 0 - 17.5 µl of molecular grade water, making a total reaction volume of 20 µl. Incubation was carried out at 37˚C for 30 minutes followed by enzyme deactivation at 80˚C for 20 minutes. Preceding long PCR amplification, restriction digestion can selectively inhibit the amplification of closely related DNA sequence families, thus, enabling the targeted amplification of one member of a group of highly homologous sequences (Her and Weinshilboum, 1999).

2.7.2 PCR amplification

The envelope gene V1-V2 region of the extracted gDNA (**Figure 2.4**) was amplified according to the method described by Abrahams *et al.* with modification (Abrahams *et al.*, 2019). First round PCR/ PCR Round 1 (R1) for the V1-V2 region was carried out using KAPA2G Fast Multiplex Mix (Roche, Basel, Switzerland) whereby each reaction contained 1 µl of forward primer (V1F_UCT) at 10 µM concentration, 1 µl of reverse primer (V2R_PID_UCT) at 10 µM concentration, 25 µl of KAPA2G Fast Multiplex Mix (2x) at 1x concentration and 20 µl of digested gDNA. For the amplification of DolPHIN-2 study samples, the integrase gene was also included. Hence, 1 µl of forward primer (C9dn_UP_UCT) and 1 µl of reverse primer (C9dn_PID_UCT) were added to the mix (**Table A12**, Appendices). To make a total volume of 50 µl, 1 - 3 µl of molecular grade water (mg H_2O) were included (**Tables A17 and A18**, Appendices).

Amplification was carried out under the following cycling conditions: initial denaturation at 95˚C for 5 minutes followed by 10 cycles each consisting of further denaturation for 15 seconds at 95˚C, 30 seconds annealing at 65˚C and 45 seconds extension at 72˚C. These 10 cycles were repeated twice, with lowered annealing temperatures of 60˚C and 58˚C, respectively. Final extension was then carried out for 1 minute at 72˚C followed by a 4˚C hold.

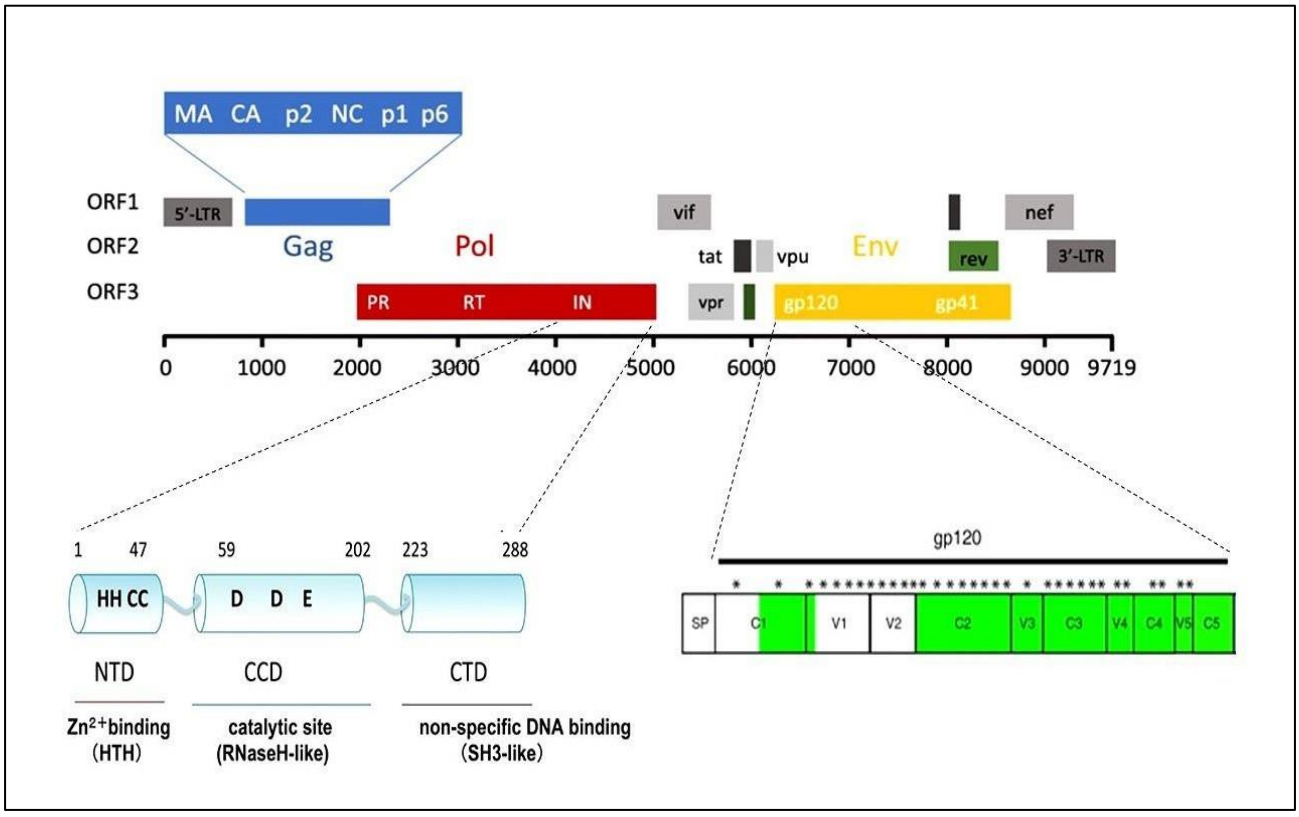

Figure 2.4. Diagram illustrating HIV-1 *env* V1-V2 and *pol* integrase regions (Caffrey, 2011; Cervera *et al.*, 2019; Masuda, 2011) amplified using the Illlumina MiSeq platform.

2.7.3 Indexing PCR

For second round/ round 2 PCR (R2), Illumina MiSeq indices were added to the first-round amplicons by PCR. Mastermixes were prepared as two separate mixes. Index mix 1 consisted of 1 µl of dNTPs (Thermo Scientific, Massachusetts, USA) at 10 µM concentration, 1 µl of universal adaptor (reverse primer) at 10 µM concentration, 3.5 - 7.5 µl water as well as 1 µl of index adaptor (forward primer) at 10 µM concentration per reaction. Unique index adaptors were used for each independent sample. Index mix 2 consisted of 9.63 µl water, 2.5 µl Expand High Fidelity PCR buffer (Roche diagnostics, Basel Switzerland) at 10x concentration and 0.38 µl Expand High Fidelity PCR enzyme (Roche diagnostics, Basel, Switzerland) at 1x per reaction, of which 12.5 µl were added to each index mix 1 reaction, followed by 2 - 6 µl of bead-purified first round PCR product, making a total volume of 25 µl per reaction (section 3.2.2, Appendices).

The mixture was amplified as follows: initial denaturation for 2 minutes at 94˚C, 35 cycles each consisting of 15 seconds at 94˚C of further denaturation followed by 30 seconds of annealing at 63˚C and extension at 72˚C for 45 seconds plus an additional 10 seconds for each cycle. That was followed by 7 minutes of further extension at 72˚C and infinite hold at 4˚C.

2.7.4 Bead-based PCR product purification

Post PCR products were purified using AMPure XP beads (Beckman Coulter, California, US) in a dedicated hood. To each reaction, a bead-to-PCR product volume ratio of 0.7 was used and the mixture incubated at room temperature for 5 minutes before being placed on a magnet for another 5 minutes following which the liquid was aspirated and discarded. Two bead washes were performed with 200 µl of fresh 70% ethanol with a 30 second incubation time following each addition. The beads were air dried and resuspended in 25 µl of molecular grade water off-magnet. Following incubation at room temperature for 1 minute, the tubes were placed back on the magnet and incubated for 3 minutes. The eluent was aspired and transferred to 1.5 ml low- bind tubes.

2.7.5 Agarose gel electrophoresis

To verify whether DNA fragments of the right size were produced following PCR, PCR products were run on a 1.2% w/v agarose gel. Agarose gels were made with 1x Tris-Borate EDTA (TBE) buffer. Loading dye (6x) (Thermo Scientific, Massachusetts, USA) containing GelRed™ (Biotium, California, USA) was added to all DNA samples to a 1x final concentration before electrophoresis. GelRed™ allows for the visualisation of DNA molecules on an agarose gel as it is an intercalating agent that interacts with DNA and fluoresces upon exposure to ultraviolet (UV) light. A molecular weight marker of 1 kb (New England Biolabs, Massachusetts, USA) was included. Electrophoresis was performed at 100 V for between 60 and 75 minutes. Gels were visualised under UV light using a Gel Doc™ XR+ Gel Documentation System (BIO-RAD, California, USA).

2.8 Sanger sequencing

env V1-V2 amplicons from selected 8E5 spiked breastmilk samples and plasma control samples from the CAPRISA 002 cohort were bead-purified following indexing as per the method described in section 2.7.3 and screened using Sanger sequencing at the Stellenbosch University Central Analytical Facility (CAF) using an ABI 3000 Genetic Analyser (Applied Biosystems, Foster City, California, USA) and BigDye terminator reagents.

2.9 Illumina Miseq Next Generation Sequencing Library Preparation

2.9.1 Gel purification of PCR amplicons

For Illumina Miseq, 25 µl of indexed PCR products was mixed with 5 µl of loading dye containing GelGreen™ (1x) (Biotium, CA, USA) and loaded alongside 4 µl of a 100 bp DNA ladder (New England Biolabs, Massachusetts, USA) , also containing GelGreen™ loading dye, and ran on a 1.2% agarose gel for 45 - 60 minutes at 70 – 100 V. A blue transilluminator was used to view and excise target DNA bands (between 500 bp and 1 kb) using a scalpel and weighed on a bench scale.

The QIAGEN MinElute gel extraction kit (QIAGEN, Hilden, Germany) was used, following the manufacturer's protocol. The kit consists of buffer QG which solubilize the excised agarose gel slice and provide the appropriate conditions for DNA to bind to the silica membrane in MinElute columns. The latter adsorb DNA in the presence of high concentrations of salt while contaminants pass through the column. Binding buffers in the MinElute Spin Kits provide the correct salt concentration and pH for adsorption of DNA to the MinElute membrane while Buffer PE wash away unbound primers and impurities such as salts, agarose and dyes. Pure DNA was then eluted in 12 µl of buffer EB (QIAGEN, 2020).

2.9.2 Pooling

DNA samples were split into three pools based on their concentrations, namely, low (from 0 ng/µl to 10 ng/µl), mid (from 10 ng/µl to less than 19 ng/µl) and high (higher than or equal to 19 ng/µl) concentrations. Samples were pooled using equal ng amounts of each sample to ensure that each sample was equally represented. In the high concentration pool 150 ng of each sample was added to the pool, in the mid pool 50 – 100 ng was added dependant on the lowest concentration and in the low concentration pool, up to 20 ng was pooled per sample. The pool concentration after beads clean-up was quantified using the Qubit 1X dsDNA High Sensitivity assay Kit (Thermo Fisher Scientific, Waltham, US) according to manufacturer's instructions. This assay is based on the fluorescent detection of double stranded DNA bound to a DNA-binding dye. It should be noted that samples from this project were pooled with other experimental samples for efficient use of flowcells and reagents.

Each of the three DNA pools were purified using AMPure XP beads (Beckman Coulter) as previously described (section 2.7.4) with the exception of a bead-to-PCR product ratio of 0.6x and DNA was eluted in 20 µl of nuclease-free water.

2.9.3 Library preparation

Each of the pools were normalised to 4 nM concentration and 5 µl of each library was pooled together to make the Pooled Amplicon Library (PAL). Equal volume of each pool added to the library ensures equimolarity across pools. The following formulae was used to calculate molarity of the pools using the average DNA size as determined via TapeStation™ (Aligent 4200 TapeStation system):

$$nM = \frac{Qubit\ Concentration\ (ng/\mu l)}{\frac{660g}{mol}\ x\ Avg\ DNA\ size}\ x\ 10^6$$

A total of 1 µl of each of the PCR product pools were run on a Tapestation using D1000 tape and reagents. This allows assessment of DNA fragment quality and provides average DNA size used for normalization of PAL.

Library preparation was performed by Miss Lynn Tyers, Division of medical virology, Faculty of Health Sciences, UCT using a PhiX Control v3 Library (Illumina, 2013) derived from a small, well-characterized bacteriophage genome (Mukherjee *et al.*, 2015) as reference. The Illumina Miseq run was also performed by Miss Tyers on the Illumina Miseq sequencer using the Miseq V3 60°Cycle kit (Illumina, San Diego, California, USA) where 3.8 µl of 100 µM Old Nextera Primer was spiked into well 12 of the Miseq cartridge and the entire 2 pM final library was loaded into the sample well on the cartridge. The sequencing run lasted approximately 65 hours.

2.9.4 Sequence analysis

Sequences produced by Illumina Miseq were initially processed using an in-house pipeline to remove poor quality reads. A program named Motifbinner2.R (https://github.com/HIVDiversity/MotifBinner2) was run by Miss Lynn Tyers to process the raw data to create FASTQ files containing the quality-trimmed forward and reverse reads for each sample. Motifbinner2.R excluded sequences below a certain quality threshold (Labuschagne, 2018). The following analysis was performed on the srvubugal04 server. Forward and reverse reads of each sample were merged using PEAR (Zhang *et al.*, 2014) and Seqmagick (GitHub - fhcrc/seqmagick: An imagemagick-like frontend to Biopython SeqIO) was used to convert the resulting merged assembled sequences to a FASTA file. Subsequently, identical sequences were collapsed using the dereplicate function in VSEARCH (Rognes, 2016) (**Table A19**, Appendices) implemented in a custom script written by Mr. David Matten, with a 1% frequency cut-off, eliminating variants that made up less than 1% of the total number of sequences to exclude variants generated through sequencing error. In cases where 1% of the total number of sequences amounted to less than 1, a minimum replicate value of 2 was used to ensure that sequences represented less than 3 times were excluded.

An additional step was taken to exclude non-HIV sequences if the dereplicated file produced more than 1 sequence using a custom in-house pipeline developed by Dr Colin Anthony. The final sequences produced were viewed on AliView (Larsson, 2014) and their associated sequence identities were confirmed using BLASTn (Altschul *et al.,* 1990).

In addition to the above analysis, CAPRISA sample sequences underwent hypermutation screening for the presence of hypermutation generated by the APOBEC-3G host restriction enzyme, where the final sequences of each sample were uploaded to the Hypermut tool (LANL, 2014).

MEGA X (Kumar *et al.*, 2018) was used for evolutionary analysis of CAPRISA samples through the generation of a phylogeny using the neighbour joining method. The evolutionary DNA distances were computed using the Maximum Composite Likelihood method and given in the unit of the number of base substitutions per site. All positions with gaps were removed for each sequence pair (pairwise deletion option).

Chapter 3: Results

3.1 Optimization of the droplet digital PCR method for HIV-1 quantification from 8E5 cells

To evaluate whether the probe-based ddPCR technique was effective at detecting and quantifying HIV genomes, assay optimization was performed using 8E5 cell line gDNA over a range of concentrations. We specifically evaluated the sensitivity of the system to quantify HIV-1 *gag* and *pol* genes. The 8E5 cell line (ATCC 8993) contains a single integrated defective copy of subtype B HIV-1 LAV proviral DNA per cell (Folks, 1986).

3.1.1 ddPCR primer/probe compatibility with the HIV-1 LAV strain from 8E5 cells

Since the 8E5 cell line harbours subtype B HIV genomes whilst the dominant HIV strain in South Africa is subtype C, it was important to evaluate the detection of 8E5 HIV DNA using existing ddPCR primers and probes in our setting. To assess whether existing ddPCR HIV subtype C *gag* and *pol* primers and probes from the HIV Diversity Group (**Table A4**, Appendices) could effectively amplify 8E5 (subtype B) HIV DNA, the primers and probes were aligned with the HIV-1/LAV/LAI genome reference sequence (Genbank accession # A04321) and assessed for compatibility.

Nucleotide mismatches were observed between the subtype C and reference subtype B sequence in the *gag* forward primer, Gag Fwd, and the *gag* probe in the 5' region and in the middle of the sequence, respectively (**Figure 3.1**). However, the *gag* reverse primer, Gag Rev, showed 100% match to the HIV subtype B sequence. A single mismatch was observed in the *pol* reverse primer, Pol Rev 4, at the 3' end. Two and one mismatches were found in the middle of the *pol* probe and *pol* forward primer, Pol Fwd 6, sequences, respectively. Due to multiple mismatches observed with the subtype C *pol* primers and probes, including in the 3' end, the latter were not considered for amplification. In contrast, subt ype B ddPCR *pol* primers and probes reported by Strain *et al*. each demonstrated 100% match with the HIV subtype B sequence. Thus, going forward, subtype C *gag* primers and probes were used along with subtype B *pol* primers and probes reported by Strain *et al.* (2013) (**Table A4**, Appendices).

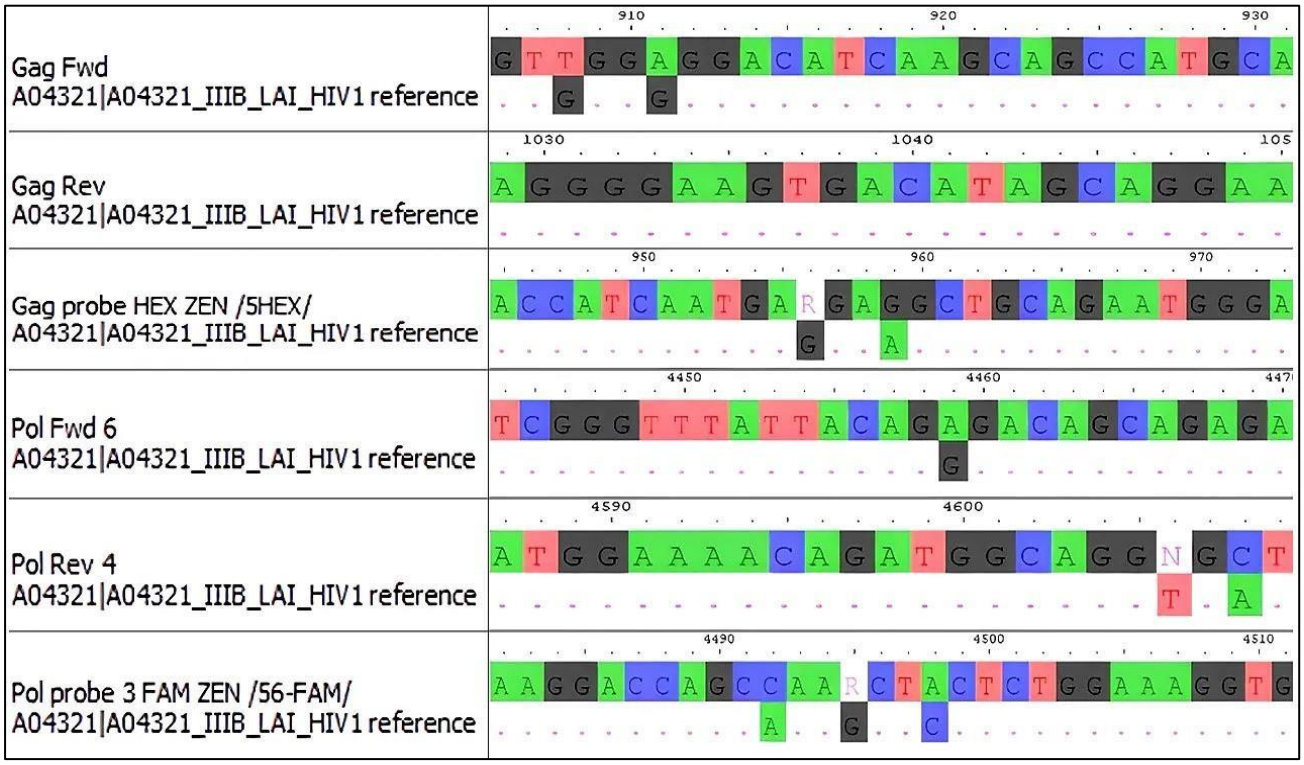

Figure 3.1. HIV-1 subtype C *gag* and *pol* primers and probes demonstrating nucleotide mismatches to the reference 8E5 HIV-1/LAI/LAV genome. Dots represent identity. Primers are shown in the 5' to 3' orientation.

Additionally, 2-LTR circles representing a portion of unintegrated HIV genomes in the cells of the host are of medical relevance (see section 1.2, Chapter 1). Hence, 2-LTR circle primers and probe were designed and synthesized for ddPCR based on previously reported subtype B counterparts (Strain *et al.*, 2013). Aligning of the subtype B LTR primers and probe locations to HIV subtype C database sequences from the Sequence Locator tool (LANL, 2019) demonstrated good conservation in the forward and reverse primers and limited diversity in the probe sequence as shown by the logograms illustrated in **Figure 3.2**. Subsequently, the HIV subtype C LTR primers and probe sequences obtained were analysed for compatibility in terms of melting temperatures, GC content, delta G values and secondary structures using the OligoAnalyzer™ tool (IDT, 2024). Results shown in **Table 3.1** demonstrate that the 2-LTR circle primers and probe were compatible with GC content ranging from 41.7% to 48.6% and mean melting temperatures ranging from 56.2°C to 60.8°C. The number of possible hairpin

structures ranged from 1 - 4, with delta G values ranging from -0.84 to 0.09 kcal/mol. Delta G of self-dimers ranged from -10.23 to -0.96 kcal/mol.

The HIV subtype C -2 LTR primers and probes had similar melting temperatures (within 5˚C of each other) and lower than the published PCR annealing temperature (Strain *et al*., 2013), hence preventing the formation of hairpin structures, which would reduce reaction efficacy. A probe melting temperature of 60.8˚C, higher than that of the primers, should ensure that the probe binds to the target first, before replication by Taq polymerase (Whitham, 2012). The CG content was within 35 – 65 % as recommended for primer binding (Prediger, 2013). The Delta G values of hairpin structures and self-dimers were higher than -9.0 kcal/mol except for one probe self-dimer (-10.23 kcal/mol), indicating low possibility of secondary structure formation (IDT, 2024).

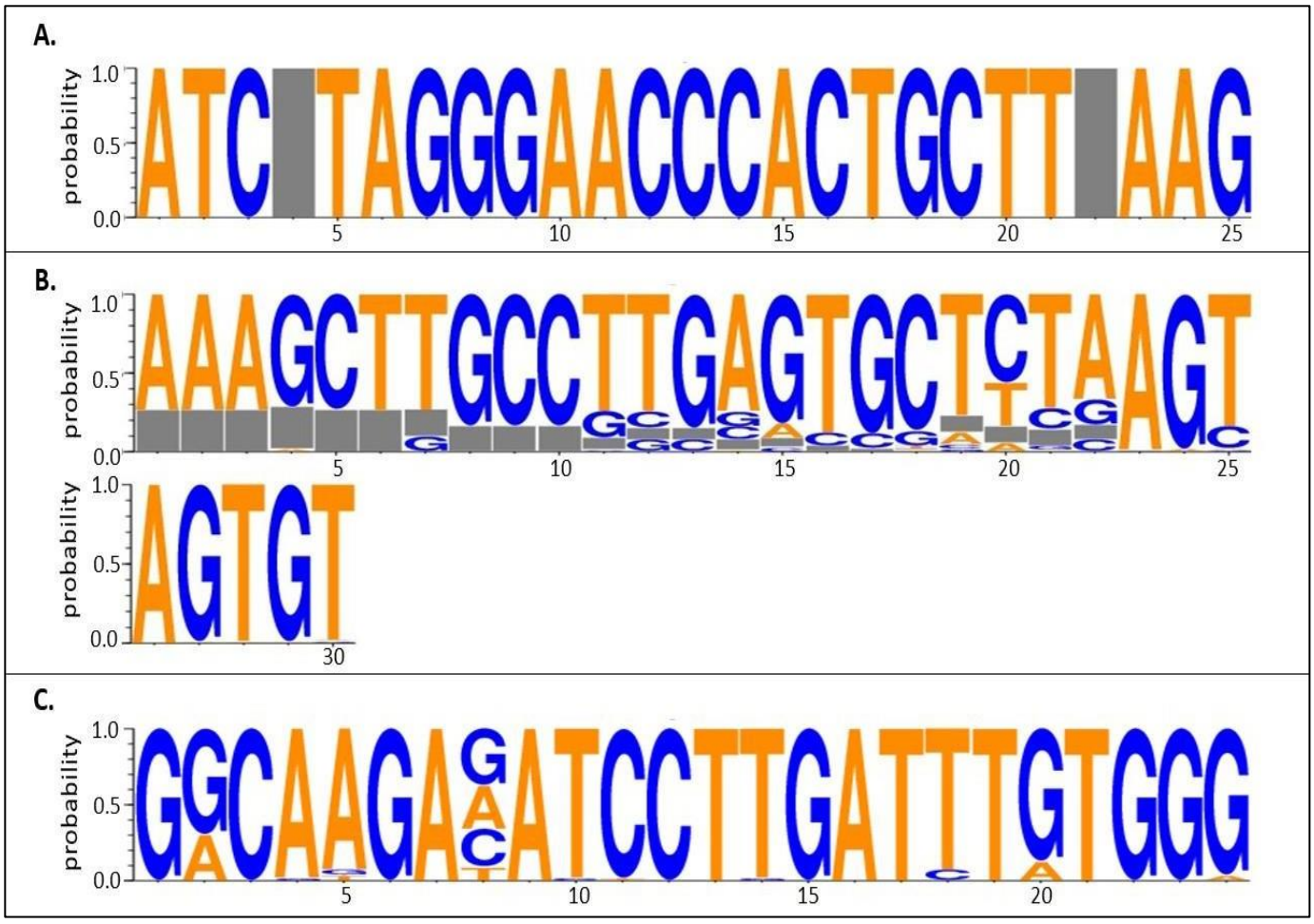

Figure 3.2. Logograms of HIV subtype C sequence alignments representing LTR primers and probe sequences. **A**. Illustrates LTR forward primer, **B**. Illustrates LTR probe, **C**. Illustrates LTR reverse primer. Nucleotide position is shown on the x-axis and the y-axis provides the probability of occurrence of each nucleotide across a sequence set. The grey bars represent gaps in the HIV subtype C consensus sequence due to variability amongst individuals (problematic sequences were removed). Primers are shown in the 5' to 3' orientation.

In cases where multiple nucleotides were observed at a nucleotide site, the one present in highest proportion was chosen for the primer or probe sequence. When multiple nucleotides were present in equal/similar proportions, the IUPAC degenerate base code was used to represent them all.

Table 3.1. Properties of HIV subtype C LTR primers and probes designed.

LTR Primer ID	Sequence	HXB2 positions (from genome start)	% GC content	Mean melting temperature ˚C
2-LTR circle forward_sub C	ATC TAG GGA ACC CAC TGC TTA AG	5' LTR: 500- 552 3' LTR: 9586- 9607	47.8	56.2
2-LTR reverse_sub C	CCC ACA AAT CAA GGA TBT CTT GCC	5' LTR: 28- 51 3' LTR: 9113- 9136	48.6	57.8
2-LTR probe HEX	/5HEX/AC ACT ACT T/ZEN/A RAG CAC TCA AGG CAA GCT TT/3IABkFQ/	5' LTR: 530- 559 3' LTR: 9615- 9644	41.7	60.8

3.1.2 Testing HIV-1 provirus quantification from 8E5 cells

To assess whether the HIV-1 proviral genome present in 8E5 cells can be effectively quantified by the ddPCR platform, 8E5 cells at a range of concentrations were prepared for amplification. The UCT HIV diversity group ddPCR protocol, SOP Nr: HDL_P3.0SOP44.1 was followed. Compatible primers and probes (see section 3.1.1) based on those published by Strain *et al.* (Strain *et al.*, 2013) were used to amplify *pol* (HXB2 positions 2536 – 2662) and HIV-1 *gag* (HXB2 positions 1360 - 1504) regions. An RPP30 (RNAse P) primer/probe set (**Table A4**, Appendices) was also used for host gDNA quantification.

gDNA extracts from 8E5 cells were diluted from a starting concentration of 10 ng/µl following a 5-fold serial dilution down to 0.08 ng/µl and a volume of 5 µl of dilution product was used per ddPCR reaction. Multiplexing *gag* and *pol* probes and primers resulted in fluorescence signal at two different amplitudes per channel for both positive and negative droplets (**Figure 3.3**). For instance, on a Quantasoft™ 1-D plot, in the FAM amplitude channel (*pol,* blue), positive droplet clouds were observed at amplitudes of 6700 and 8600 relative fluorescence units (RFU) (**Figure 3.3 A**). In the HEX amplitude channel (*gag,* green), positive droplet clouds were observed at approximate amplitudes of 2400 and 2900 RFU while negative droplet clouds (grey) were at 950 and 1400 RFU amplitudes (**Figure 3.3 B**). This was especially observed in wells containing higher DNA concentration inputs as shown in the form of double amplitude bands (blue and green) in **Figures 3.3 A and B** (separated by a black line). This rendered the HIV copy number quantification less reliable.

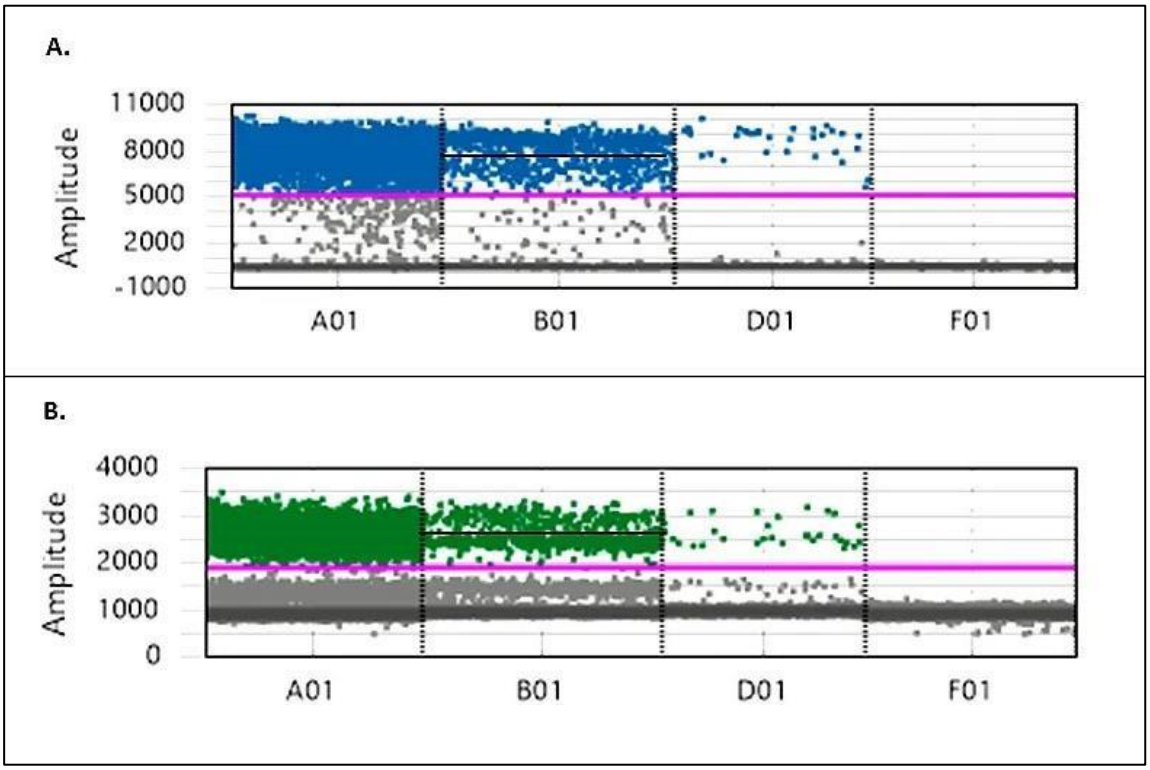

Figure 3.3. QuantaSoft™ 1-D amplitude plots of HIV *pol* (A, FAM) and *gag* (B, HEX) amplification signal (positive droplets) following a 5-fold serial dilution from a starting concentration of 10 ng/µl to 0.08 ng/µl. These are illustrated in wells A01, B01 and D01 respectively (wells containing less than 9000 droplets were excluded). Wells F01 contain a water non-template control. The black lines in B01 wells indicate separation of two positive amplitude droplet clouds. **A**. *pol* positive droplets are illustrated in blue while **B**. *gag* positive droplets are illustrated in green. The grey amplitude clouds represent negative droplets. The amplitude thresholds for exclusion of negative/background signals were set at 4910 for *pol* and 1886 for *gag* (pink lines).

Thereafter, multiplexing experiments were carried out with housekeeping gene *rpp*30 and *pol* primers and probes to investigate whether probe cross reactivity between HIV-specific probes was responsible for the dual amplitude signal. When *rpp30* and *pol* probes and primers were combined, two positive amplitude clouds were observed for *pol* (blue) as well as *pol* and *rpp30* double positive (orange) as illustrated in **Figure 3.4** instead of the expected single positive clouds. It was therefore determined that *pol* amplification would be carried out in a singleplex experiment, independently of *rpp30* and *gag*, to avoid the signal problems

observed with multiplexing.

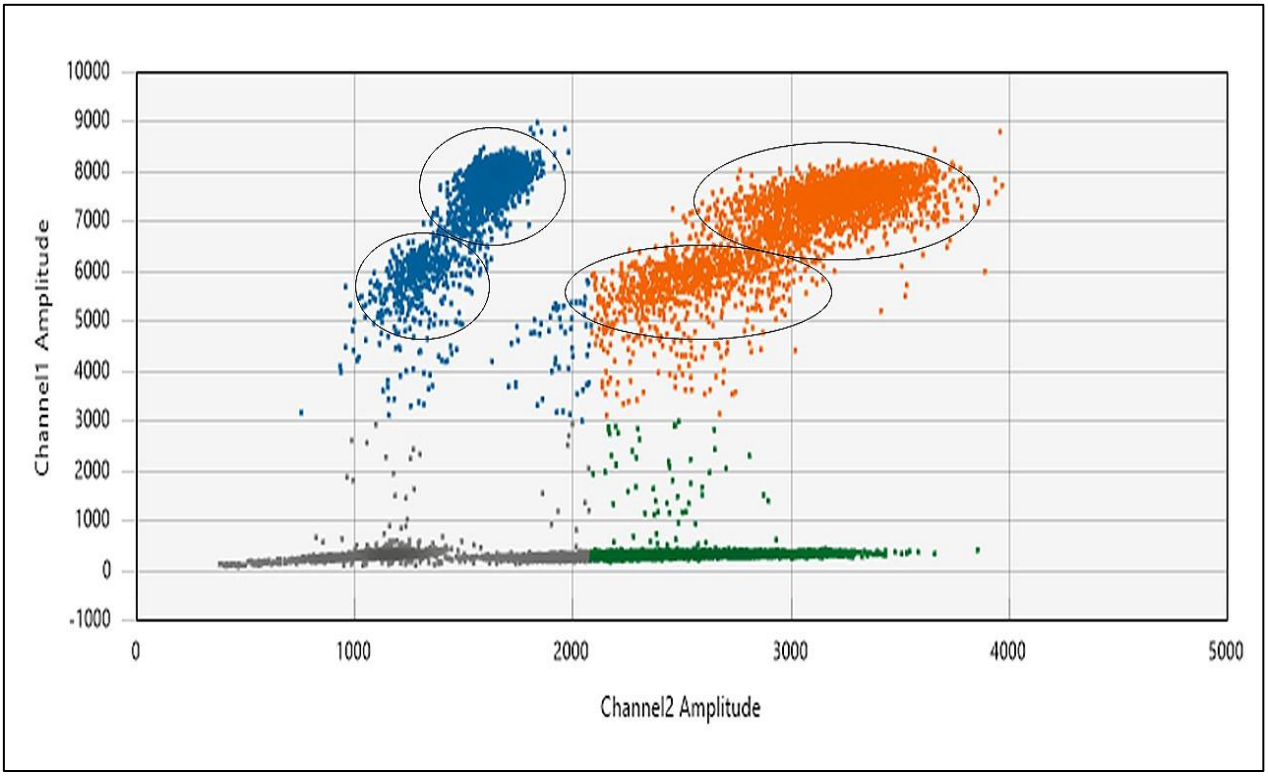

Figure 3.4. QuantaSoft 2-D amplitude plots depicting signal for multiple wells, with amplification of HIV *pol* and *rpp30* genes, following a 5-fold serial dilution from a starting concentration of 20 ng/µl to 0.032 ng/µl. Wells were multiplexed with both *pol* and *rpp30* primers and probes where the orange cloud represents double positive droplets while grey amplitude clouds represent double negative droplets. Blue droplets represent *pol*-only positive droplets while the green cloud represent *rpp30*-only positive droplets. Wells containing less than 9000 droplets were excluded.

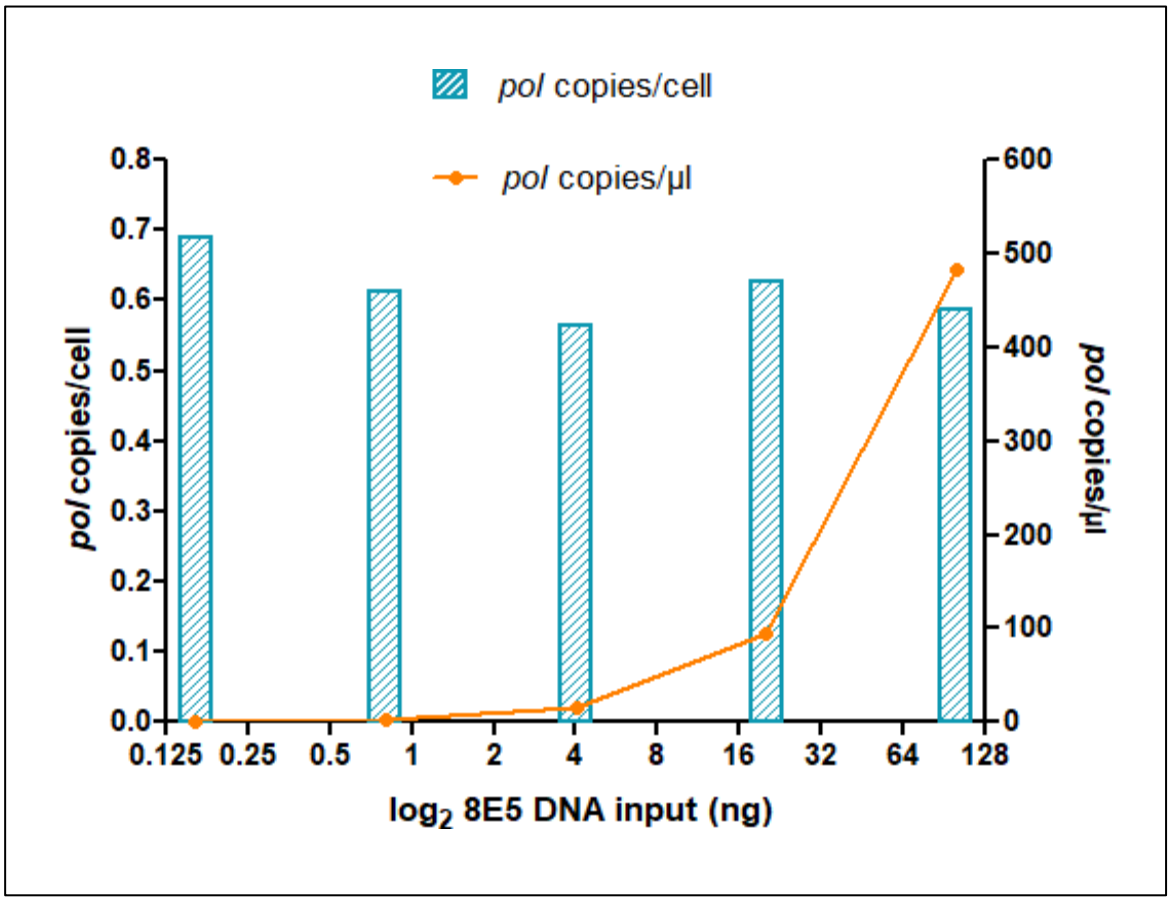

Figure 3.5. Bar chart illustrating 8E5 cell line *pol* copies/cell on the left y-axis and *pol* copies/µl on the right y-axis plotted against $\log_2$ 8E5 genomic DNA inputs (ranging from 100 ng to 0.16 ng following a 5-fold dilution series) on the x-axis. The number of *pol* copies per 8E5 cell across different 8E5 concentrations is illustrated in blue bars while the orange dots connected by a line represent *pol* copies/µl for each concentration.

The HIV concentration in *pol* copies/µl reflect the 5-fold serial dilution as illustrated by the line graph (**Figure 3.5**). Across all dilution concentrations, *pol* copies/cell were similar, as expected from ddPCR, indicating that the platform is accurate at the abovementioned concentrations and is therefore sensitive for HIV DNA quantification from 8E5 cells.

All future ddPCR experiments were singleplexed.

3.2 Spiked-in HIV in uninfected breastmilk samples

Objective one sought to evaluate the sensitivity of HIV DNA copy number detection using ddPCR on HIV negative breastmilk samples spiked with cells infected by HIV-1 defective genome. For this purpose, HIV uninfected breastmilk samples obtained from the Mowbray breastmilk bank in Cape Town were spiked with varying combinations and concentrations of 8E5 cells as well as background cells. For optimal HIV detection (and quantification in cells from individuals on ART) the ddPCR protocol by Strain *et al*. used as a basis for this study, recommends an input of 1000 ng of gDNA per reaction (Strain *et al*., 2013).

3.2.1 DNA recovery from HIV uninfected breastmilk spiked with 8E5 cells and PBMCs.

In order to determine the quantity and quality of gDNA recovered from breastmilk samples, following 8E5 spike-in only, gDNA was extracted from 1 million total cells for each of breastmilk donor samples 1 and 2 (BM1 and BM2) and quantified using a NanoDrop 2000 Spectrophotometer. Spiked milk samples A-E reflect the 10-fold serial dilution of 8E5 cells described in section 2.4.1 in Methods where samples A, B, C, D and E were spiked with 10 000, 1000, 100, 10 and 1 8E5 cell(s) respectively (**Table A1**, Appendices) while sample F were negative non-spiked controls.

Breastmilk spiked with only 8E5 cells produced gDNA concentrations ranging from 0.16 ng/µl to 14.62 ng/µl and 5.69 ng/µl to 10.23 ng/µl for samples BM1 and BM2, respectively in a final volume of 30 µl. Overall, these samples had a median gDNA concentration of 7.10 ng/µl and an average of 7.28 ng/µl. Their average 260/230 and 260/280 ratios were 0.23 and 0.87, respectively, both below the recommended 2.0 for 260/230 ratios and 1.8 for 260/280 ratios (**Table A21**, Appendices) which may indicate the presence of contaminants from chemicals used during the extraction process or residual proteins. The concentration of DNA was also unexpectedly low relative to the expected yield for the cell count. According to the manufacturers of the DNA extraction kit used (QIAGEN AllPrep DNA/RNA Mini kit manual), from 1 million cells, HIH/3T3, HeLa or Jurkat and Cos-7 cells produce on average 8, 6 and 7 µg

of DNA yield, respectively (QIAGEN, 2020). The ratio of DNA recovered to chemicals used may therefore have been very low, thus resulting in low 260/280 and 260/230 ratios.

Therefore, further spike-in experiments with 8E5 cells supplemented with PMBCs were performed to increase yields obtained from gDNA extraction. Two additional breastmilk samples were collected (BM4 and BM6) and spiked with the same quantities of 8E5 cells as previously described and topped up with PBMCs, making up a total of 1 million cells in each sample (**Table A2**, Appendices) as described in section 2.4.2 in Methods.

Following DNA extraction, the gDNA concentrations ranged from 25.12 - 52.45 ng/µl for BM4 and from 51.70 - 160.85 ng/µl for BM6 (**Table A22**, Appendices). Overall, spiking breastmilk samples with 8E5 cells and PBMCs increased the median and average gDNA yield to 42.57 ng/µl and 60.60 ng/µl, respectively. These concentrations were respectively 6 and 8 -folds higher as compared to 8E5-only spiked samples BM1 and BM2. With an average of 1.63 and 1.71, respectively, the 260/230 and 260/280 ratios also improved as compared to BM1 and BM2 (**Table A21**, Appendices) despite some samples still having 260/230 and 260/280 ratios below the recommended 2.0 and 1.8 respectively (Thermo Fisher Scientific, no date).

3.2.2 DNA recovery from HIV uninfected breastmilk spiked with 8E5 cells and SUP-T1 cells

In order to further increase yield from gDNA extraction, we increased the total spike-in cell number to 5 million using SUP-T1 cells (a cell line derived from T-cell Lymphoblastic Lymphoma malignant cells) (Smith *et al.*, 1988) instead of PBMCs. HIV-1 naïve breastmilk samples were spiked with 8E5 cells similarly to section 2.4.1 with some changes as described in section 2.4.3 in Methods. For instance, three distinct donor samples were pooled to produce sample BM8. Thereafter, milk samples A-E were spiked with 8E5 cells from concentrations of 50 000 - 5 8E5 cells/million cells respectively, following a 10-fold dilution series dilution. SUP-T1 cells were also spiked in the breastmilk samples such that both cell types added up to a total of 5 million cells in each sample.

Genomic DNA concentrations of pooled sample BM8 ranged from 498.99 ng/µl to 757.28 ng/µl (**Table A24**, Appendices). Spike-in with a total of 5 million cells significantly increased

DNA yield as compared to spike-in with 1 million cells total. With an average of 1.92 and 2.07, the 260/280 and 260/230 ratios above the recommended 1.8 and 2.0, respectively (Thermo Fisher Scientific, no date), indicating the absence of contaminants. The overall median and average gDNA yields were 614.58 and 631.32 ng/µl, respectively.

3.3 ddPCR of HIV from 8E5 spiked breastmilk samples

3.3.1 ddPCR of HIV-1 from breastmilk spiked with 8E5 cells and PBMCs.

To assess the sensitivity of the ddPCR platform to quantify HIV-1 DNA from a total of 1 million cells spike-in, breastmilk samples BM4 and BM6 were spiked with 8E5 cells ranging from 10 000 8E5 cells/million to 1 8E5 cell/million cells from samples A-E whilst maintaining a constant of 1 million cells total in each sample with PBMCs. ddPCR quantification of extracted gDNA was performed with *pol* and *rpp30* primers and probes, in singleplexed experiments (**Figure A3**, Appendices).

The maximum possible amount of DNA was used for each sample in wells containing *pol* primers and probes, however not exceeding 1000 ng. For BM4, DNA input ranged from 83 ng to 274 ng with an average of 194 ng. DNA input ranged from 424 ng to 1000 ng with an average of 681 ng for BM6. In wells containing *rpp30* primers and probes, DNA input for BM6 was 100 ng per reaction, as per the recommended optimal amount by Strain *et al.* (2013) and ranged from 50.9 ng to 100 ng for BM4, with an average of 83.9 ng (**Table A23**, Appendices).

According to **Figure 3.6 A**, sample A, spiked with the highest number of 8E5 cells (10 000 8E5 cells/million) resulted in the highest concentration of *pol* copies being detected, 2490 copies/million cells for BM4A and 2261 copies/million cells for BM6A (**Table A23**, Appendices). In fact, post exclusion of reaction wells due to low droplet numbers and copy numbers below the Poisson confidence interval maximum values, the only samples with average values above background levels were samples BM4A and BM6A. Moreover, the 8E5 gDNA positive control reported an average of 792 555.1 *pol* copies/million cells.

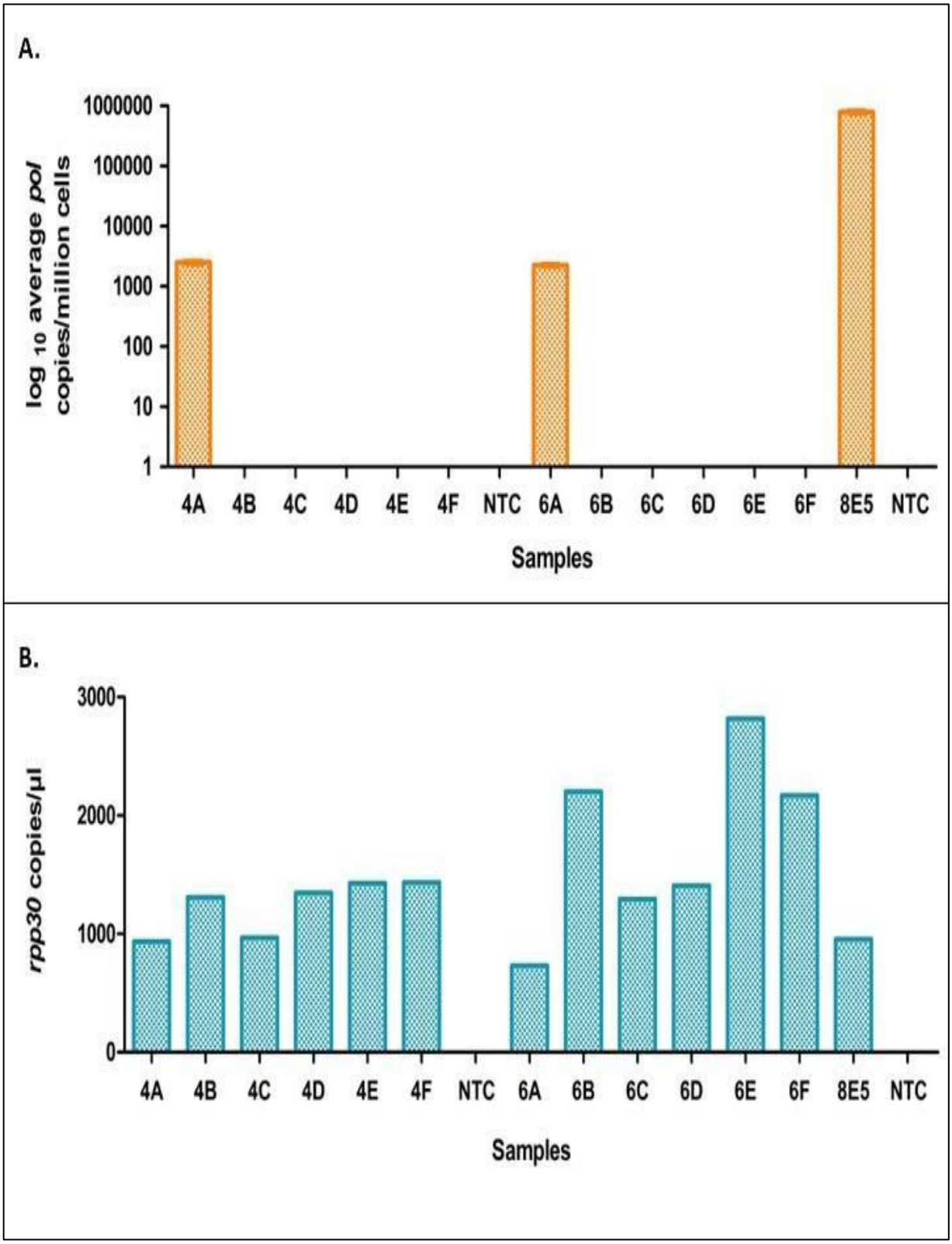
A.
log 10 average pol copies/million cells
1000000
100000
10000
1000
100
10
1
4A 4B 4C 4D 4E 4F NTC 6A 6B 6C 6D 6E 6F 8E5 NTC
Samples
B.
rpp30 copies/µl
3000
2000
1000
0
4A 4B 4C 4D 4E 4F NTC 6A 6B 6C 6D 6E 6F 8E5 NTC
Samples

Figure 3.6. Bar charts depicting *pol* (with standard deviation) and *rpp30* copies obtained from ddPCR quantification of breastmilk samples BM4 and BM6 spiked with 8E5 cells and PBMCs as well as their positive (8E5) and non-template controls (NTC). **A**. Orange bars illustrating $\log_{10}$ average *pol* copies per million cells where only samples BM4A, BM6A and the 8E5 control show detectable *pol* signal. **B**. Blue bars illustrating *rpp30* copies/µl (only one replicate used) normalized for 100 ng of DNA input.

From **Figure 3.6 B**, *rpp30* concentrations were relatively uniform across BM4 samples as they fell within a smaller range from 1434 to 690 copies/µl with a median of 950.4 copies/µl in comparison to that of BM6 which showed large fluctuations ranging from 2810 to 731 copies/µl, with a median of 1408 copies/µl, thus demonstrating non-uniformity for a given gDNA input. Additionally, the 8E5 positive control contained 954 *rpp30* copies/µl which is lower than the expected 1000 to 2000 copies/µl range (Abrahams, 2024, personal communication). Interestingly, we recorded 0.8 HIV copies/8E5 cell from the 8E5 gDNA positive control (see **Table A23**, Appendices) instead of the expected one HIV copy per 8E5 cell (Désiré *et al.*, 2001).

Overall, low gDNA concentrations and positive droplet values below background levels resulted in insufficient sensitivity for further analysis.

3.3.2 ddPCR of HIV-1 from breastmilk spiked with 8E5 cells and SUP-T1 cells.

To further explore the sensitivity of HIV detection in breastmilk, pooled breastmilk sample BM8 was spiked with 8E5 cells ranging from 250 000 to 25 cells from samples A-E (**Figure 2.3**, Methods), making up a total of 5 million cells with SUP-T1s (total cell number as used in standard PBMC-based protocols reported by Strain *et al.* (2013). This represented a conservative spike-in range of 50 000 - 5 8E5 cells/million to account for variation in per cell HIV copy number in the 8E5 cell line as seen in previous experiments in this study (see sections 2.4.3 and 3.3.1) and previously reported (Busby *et al.*, 2017). ddPCR quantification was performed with *gag* instead of *pol* primers and probes due to primer availability and *rpp30*

primers and probes as previously, in singleplexed experiments. As shown in section 3.1, *gag* probes were found to be compatible with the HIV-1 LAV strain. Regarding each sample, 1000 ng gDNA input was used in wells containing *gag* primers and probes while 100 ng was used in those containing *rpp30* primers and probes. For both gene regions, gDNA from 8E5 cells at 100 ng was used as a positive control and water as NTCs were included. Wells containing *rpp30* primers and probes were duplicated and those containing *gag* primers and probes were run in triplicate (**Figure A5**, Appendices).

As illustrated in **Figure A5** in the Appendices, sample BM8 produced higher HIV copy number concentrations overall, as compared to the previous spike-in experiments. Sample BM8F, a non-spiked breastmilk control and a single water NTC, respectively contained a few positive droplets (0.1 *gag* copies/µl each). However, these translated to DNA copy number values below the Poisson confidence interval maximum value of 0.25 copies/µl obtained for NTCs and can thus be considered as background signal. False positive droplets have also been reported as common with ddPCR (Strain *et al.,* 2013). The 8E5 positive controls produced an average of 0.4 *gag* copies/cell (**Table A25**, Appendices).

Breastmilk sample BM8E HIV copy number concentrations fell below the Poisson confidence interval maximum concentration of 0.25 copies/µl and therefore was not considered as positive. Thus, the average detectable *gag* concentration ranged from 0.37 - 72.13 copies/µl, corresponding to samples BM8D and BM8A spiked with 50 and 50 000 8E5 cells/million cells, respectively. This corresponded to an intracellular DNA copy concentration ranging from 44.29 – 9239.97 HIV copies/million cells (**Figure 3.7 A**, **Table A25**, Appendices). The lowest spike-in concentration (44.29 HIV copies/million cells) detectable by ddPCR corresponded to a limit of detection of 7 *pol* copies for this experiment.

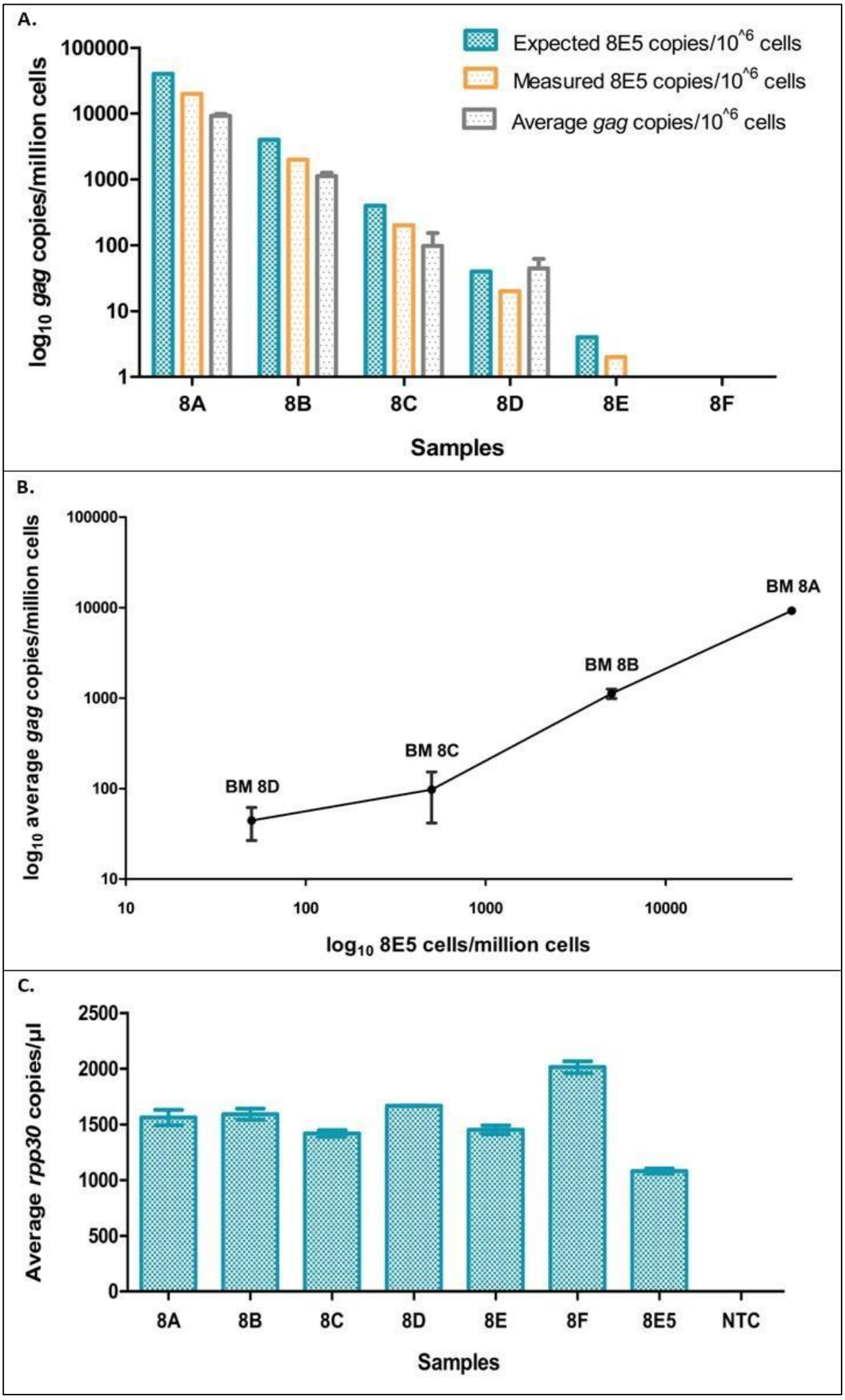
A.
Expected 8E5 copies/10^6 cells
Measured 8E5 copies/10^6 cells
Average gag copies/10^6 cells
log10 gag copies/million cells
100000
10000
1000
100
10
1
8A
8B
8C
8D
8E
8F
Samples
B.
log10 average gag copies/million cells
100000
10000
1000
100
10
BM 8A
BM 8B
BM 8C
BM 8D
10
100
1000
10000
log10 8E5 cells/million cells
C.
Average rpp30 copies/µl
2500
2000
1500
1000
500
0
8A
8B
8C
8D
8E
8F
8E5
NTC
Samples

Figure 3.7. Graphs illustrating *gag* and *rpp30* copies (with standard deviation) obtained from ddPCR quantification of breastmilk sample BM8 spiked with 8E5 cells and SUP-T1 cells, as well as positive (8E5) and negative controls (NTC). **A**. Bar chart illustrating *gag* copies/million cells from BM8 A-F spiked with 250 000, 25 000, 2500, 250, 25 and 0 8E5 cells, respectively in a total of 5 million cells per sample. Expected 8E5 HIV genome copies/million cells (blue bars) represent the number of 8E5 proviral genome copies expected per million cells for the spike-in experiment, based on the most recent prior experiment of 0.8 HIV proviral copies/8E5 cell. Measured 8E5 HIV genome copies/million cells (orange bars) represent the actual spike-in concentrations calculated based on 0.4 proviral copies/8E5 cell observed for this ddPCR experiment. Average *gag* copies/million cells (grey bars) reflect the HIV copy frequency detected from ddPCR experiments (error bars shown for 3 replicate wells). **B**. Line graph showing intracellular *gag* copies/million cells obtained from each 8E5 dilution concentration from samples BM8 A-D. **C**. Bar chart demonstrating the average *rpp30* concentration obtained from each sample.

As shown in **Figures 3.7 A and B**, from samples BM8 A- C the intracellular DNA concentration (*gag* copies/million cells) followed approximately a 10-fold dilution, reflecting the breastmilk spike-in with 8E5 cells. From sample 8D, spiked with 50 8E5 cells/million cells, accuracy and sensitivity were likely lost as the proviral concentration deviated from the 10-fold dilution series (**Table A25**, Appendices). Moreover, the 8E5 positive control showed only 0.4 HIV copies/cell as opposed to the reported 1 HIV/LAV copy per 8E5 cell for this cell line (Désiré *et al.*, 2001) or the 0.6 - 0.8 HIV copies per 8E5 cell detected in previous experiments in this study (see sections 3.1.2 and 3.3.1). It can therefore be deduced that this experiment spiked in 20 000 to 2 8E5 HIV genomes/million cells.

Therefore, HIV was detectable in breastmilk samples spiked with 20 000 to 20 8E5 proviral copies/million cells, corresponding to BM8 A-D, with signal above background levels. Yet, from BM8 A to C the measured spike-in values were higher than the ddPCR detected copy number concentration of the spiked samples, with an average 2-fold difference detected.

Additionally, illustrated by **Figure 3.7 C**, *rpp30* concentrations were relatively uniform across samples, ranging from 2013 – 1080.5 copies/µl, which is within the expected range

(Abrahams, 2024, personal communication). Therefore, increasing the total number of cells to 5 million resulted in better DNA recovery and positive signal above background levels down to 1.3$\log_{10}$ HIV copy number concentrations per million cells.

3.4 Optimization of an Illumina MiSeq Next Generation Sequencing method for sequencing of HIV in breastmilk cells

In order to optimize Illumina MiSeq NGS of HIV gDNA from breastmilk, amplification and sequencing of control gDNA samples were performed using an optimized protocol developed for RNA amplification from the HIV Diversity Group (Abrahams *et al.,* 2019). Amplification of the V1-V2 region of the HIV envelope gene (*env*) and a sub-region of the integrase gene (*int*) (see **Table A12** for HXB2 co-ordinates) was tested.

Additionally, Sanger sequencing was performed prior to Illumina sequencing to confirm that the expected region of *env* was indeed amplified.

3.4.1 Amplification of HIV *env* from 8E5 genomic DNA.

According to objective one of this study, 1 ml of HIV naïve breastmilk was spiked with 10 000, 1000, 100, 10 and 1 8E5 cells/million, respectively. To replicate this dilution series and evaluate efficacy and sensitivity of *env* amplification using primers from the Illumina MiSeq protocol, extracted 8E5 gDNA was diluted based on previously determined copy number per cell. The amount of DNA required to represent the desired number of genome copies was determined based on prior ddPCR experiments (section 3.1). To obtain the required numbers of HIV copies from 8E5 cells containing gDNA at a concentration of 70 HIV genome copies/ng, 8E5 gDNA was serially diluted 10-fold from 7.5 ng/µl to make the following concentrations: 0.75, 0.075, 0.0075 and 0.00075 ng/µl to represent 1000, 100, 10 and 1 8E5 HIV genome copies in a final volume of 20 µl, respectively (**Figure 3.8**).

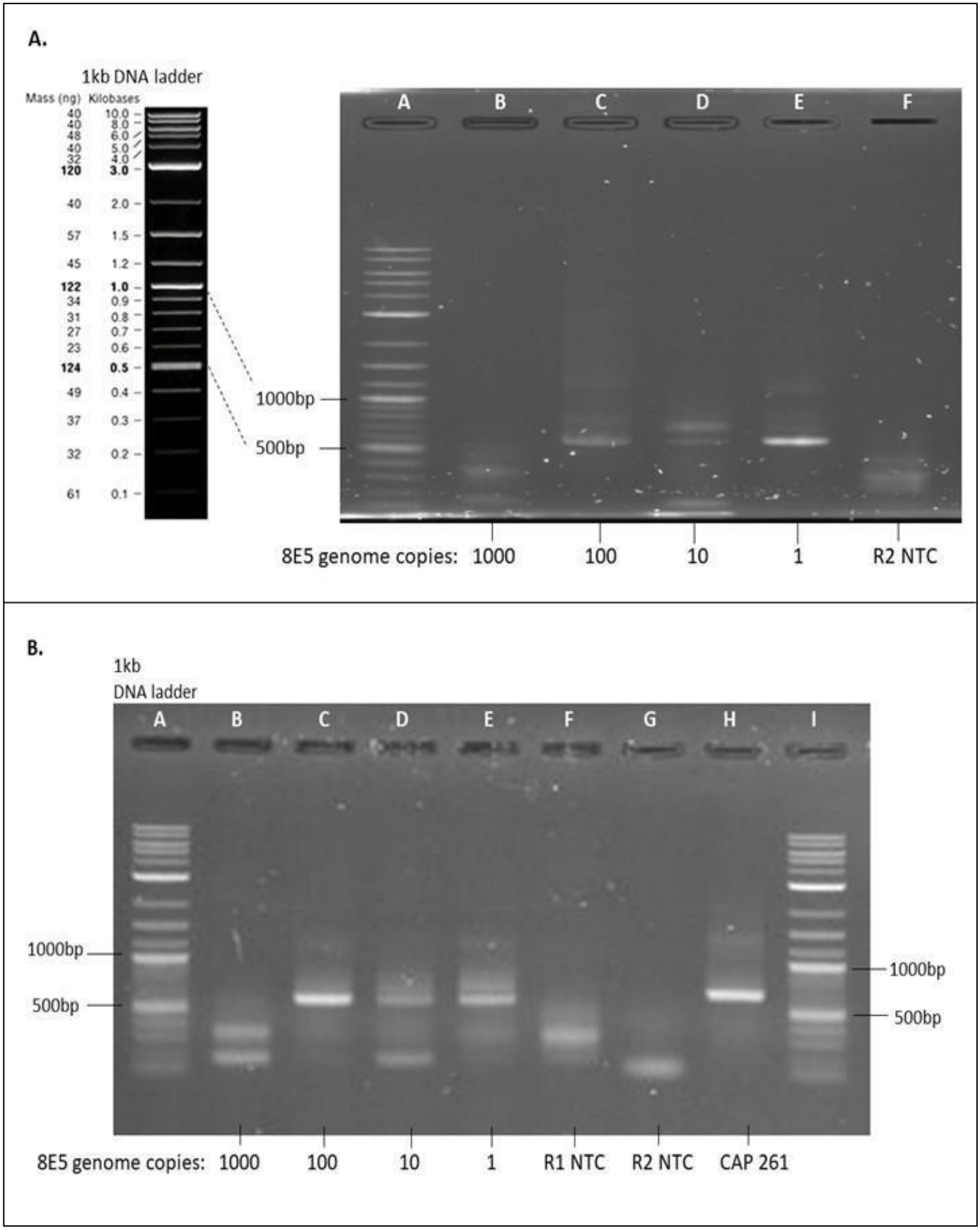
A.
1kb DNA ladder
Mass (ng) Kilobases
40 10.0
40 8.0
48 6.0
40 5.0
32 4.0
120 3.0
40 2.0
57 1.5
45 1.2
122 1.0
34 0.9
31 0.8
27 0.7
23 0.6
124 0.5
49 0.4
37 0.3
32 0.2
61 0.1
1000bp
500bp
A B C D E F
8E5 genome copies: 1000 100 10 1 R2 NTC
B.
1kb
DNA ladder
A B C D E F G H I
1000bp
500bp
1000bp
500bp
8E5 genome copies: 1000 100 10 1 R1 NTC R2 NTC CAP 261

Figure 3.8. 1% Agarose gel images following amplification of *env* V1V2 from 8E5 gDNA samples at a range of concentrations post indexing/ second round PCR. **A**. Lanes B to E contain PCR products generated from 8E5 gDNA inputs at the following concentrations: 0.75, 0.075, 0.0075 and 0.00075 ng/µl to represent 1000, 100, 10 and 1 8E5 cells, respectively. Lane F contained a PCR second round (R2) water non template control (NTC). **B**. Lanes B to E contain the same PCR products as in gel A following a 10-fold dilution of the first round (R1) PCR products. Lane F contained the indexed product from inputting 2 µl of the first round (R1) NTC to control for bands generated due to the presence of residual first round PCR primers, while Lane G contained a PCR second round (R2) NTC. Lane H contained indexed product from inputting 0.5 µl of a plasma positive control (CAP261) first round PCR product. Lanes A and I contained a 1 kb DNA ladder (Quick-Load® Purple 1 kb Plus DNA Ladder, New England BioLabs Inc.) with GelRed™ dye (Biotium), also depicted on the top right (**A**).

Following restriction enzyme digestion, *env* V1V2 first round PCR and indexing PCR, the brightest DNA band on an agarose gel was observed at the lowest concentration input of 0.0075 ng/µl in lane E of **Figure 3.8 A**, corresponding to 1 8E5 genome copy. Higher DNA input concentrations in wells B-D may have caused inhibition, thus producing unclear bands or no bands at the highest concentration input. Therefore, indexing PCR was repeated with a lowered first round DNA input volume (from 6 µl to 2 µl) and first round PCR product sample concentration was diluted 10-fold to reduce PCR inhibition as shown in **Figure 3.8 B**. The brightest DNA band similar in size to the positive control was observed in lane C containing PCR product from an input of 100 8E5 genome copies. The expected band was not obtained at the 1000 genome copies input, once again likely due to inhibition by high copy numbers. The lowest DNA input of 1 8E5 genome copy in lane E also produced a band within the expected range, thus demonstrating the sensitivity of this assay. Additionally, lanes B, D and E produced low-sized bands which correspond to that of NTC controls in lanes F and G, indicating the presence of primer dimers (Brownie *et al.*, 1997).

3.4.2 Testing of Illumina MiSeq integrase primers for next generation sequencing.

To allow us to investigate the presence of Dolutegravir-associated drug resistance mutations in proviral DNA from DolPHIN-2 study breastmilk cells samples, integrase primers designed for the HIV Diversity Group multiplexed Illumina MiSeq from RNA method were tested on 8E5 gDNA but amplification failed and it was later found to be due to mismatches in the primer binding sites (data not shown). The primers were then tested on a HIV subtype C infectious molecular clone (IMC) to amplify this region of HIV-1 *pol* from a DNA template (**Figure 3.9**). An IMC containing the HIV genome of CAPRISA 002 cohort (Kwa-Zulu Natal) participant CAP45 was selected as it represented a 100% match in the binding sites for the *integrase* primers. Although not known, all DolPHIN-2 samples were expected to be subtype C based on dominance of the subtype in South Africa.

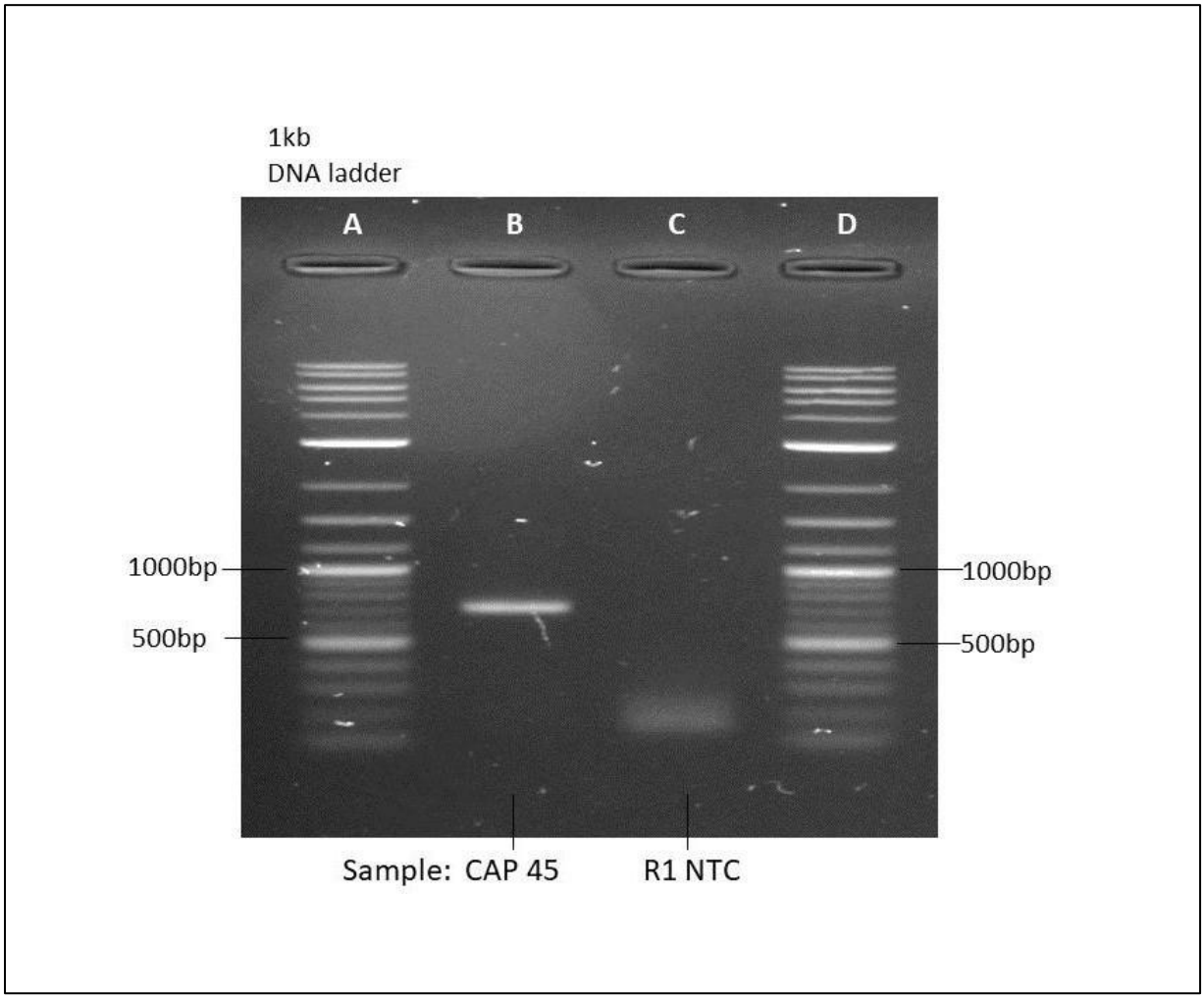

Figure 3.9. 1% Agarose gel image of CAPRISA sample CAP45 following first round PCR (R1) with *integrase* primers, shown in lane B while lane C contained a first round PCR (R1) non-template control (NTC). Lanes A and D both contained a 1kb DNA ladder (Quick-Load® Purple 1 kb Plus DNA Ladder, New England BioLabs Inc.) with GelRed™ dye (Biotium).

PCR amplification of the CAP45 IMC produced a band within the expected size range (500 - 700 bp), thus confirming that the *integrase* primers work on HIV-1 subtype C DNA samples.

3.4.3 Amplification and sequencing of HIV from genomic DNA of individuals on antiretroviral treatment

To evaluate whether our Illumina MiSeq PCR protocol tested on 8E5 gDNA was sensitive enough to amplify low-copy number HIV from individuals on antiretroviral treatment i.e., from the HIV-1 reservoir, three PBMC-extracted gDNA samples of varying proviral copy number concentrations were selected from the Centre for the AIDS Programme of Research in South Africa (CAPRISA) 002 cohort (van Loggerenberg *et al.,* 2008). The proviral copy numbers as measured by ddPCR by Dr Sherazaan Ismail (Division of Medical Virology, Faculty of Health Sciences, UCT) were selected to range from high to medium to low, relative to measures for the cohort (**Table 3.2**). The CAPRISA 002 study measured virological, immunological and clinical events in acute and early infection and over 5 years of antiretroviral treatment. CAPRISA samples CAP206, CAP244 and CAP277 had ART exposure for 4 years, 5 years and 2 years respectively. The HIV reservoir sizes of the abovementioned samples are given in **Table 3.2** below.

Table 3.2. Proviral copy number concentrations of selected on-ART CAPRISA samples.

Sample ID	CAP206	CAP244	CAP277
gag copies/million cells	905	704	82
pol copies/million cells	1309	484	92

Samples CAP206, CAP244 and CAP277 were classified as having high, medium and low viral loads, respectively. The *gag* copies/million cells ranged from 82 to 905 while *pol* copies/million cells ranged from 92 to 1309.

Each gDNA sample was obtained at a concentration of 200 ng/µl, from which 1000 ng was used for PCR amplification.

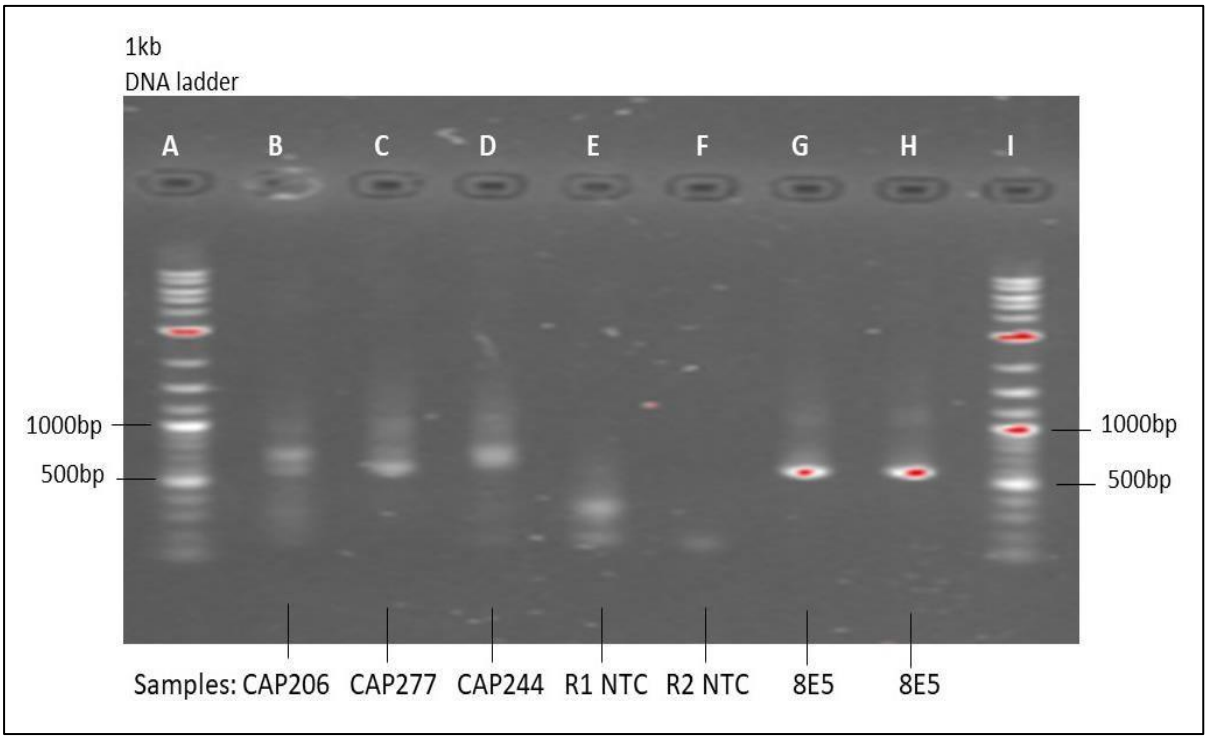

Figure 3.10. 1% Agarose gel image following *env* V1V2 PCR amplification of PBMC gDNA from the CAPRISA cohort post indexing/ second round PCR. Lanes B, C and D contained samples CAP206, CAP277 and CAP244, respectively. Lane E contained the indexed product from inputting 2 µl of the first round PCR (R1) non template control (NTC) to control for bands generated due to the presence of residual first round PCR primers, while lane F contained a PCR second round (R2) NTC. Lanes G and H each contain 0.015 ng of 8E5 as positive controls while Lanes A and I contain a 1kb DNA ladder (Quick-Load® Purple 1 kb Plus DNA Ladder, New England BioLabs Inc.) with GelRed™ dye (Biotium).

All three CAPRISA samples were successfully amplified and produced DNA bands on an agarose gel within the expected range of 500 to 1000 base pairs (**Figure 3.10**), indicating that the primers were effective at amplifying DNA from clinical samples with high to low proviral loads. CAPRISA participant CAP206 gDNA amplification (lane B) produced double bands which could be due to either non-specific amplification or target regions of different sizes in this participant. Bands in lanes E and F NTCs may be primer dimers due to their small size while G and H positive control lanes confirm that bands produced in lanes B to D correspond to the expected size.

In order to check whether the desired genome region had been PCR amplified, samples selected for Illumina Miseq sequencing were first sent for Sanger sequencing following bead-based DNA purification. All amplicons were confirmed to be *env* V1V2.

The target sequence was successfully sequenced for all three CAPRISA samples on the Illumina MiSeq platform. Prior to quality control, the raw read depth ranged from 59 105 to 211 749 reads and the mean quality score of forward sequences ranged from 32.48 to 34.53 and from 30.40 to 34.52 for reverse reads. After removal of short and low-quality reads and merging the forward and reverse reads, the number of sequences ranged from 42 382 to 146 121 (**Table 3.3**). Identical sequences were collapsed using the dereplicate function in VSEARCH (Rognes, 2016) and sequences present at less than 1% frequency post dereplication were excluded.

Table 3.3. CAPRISA samples Illumina MiSeq data outputs

Sample	DNA input/ng	Number of assembled (Q score >25) sequences	Number of sequences collapsed	Number of variant sequences after applying 1% frequency cutoff
CAP206	1000	146121	11137	3
CAP244	1000	42382	859	1
CAP277	1000	82012	7984	6

Following the 1% frequency cutoff, the number of variant sequences produces ranged from only 1 to 6. Samples CAP206 and CAP244 matched with their corresponding sequence identities using BLASTn (National Center for Biotechnology Information [NCBI], no date). Since CAP277 was not available in BLASTn databases, CAP277 sequences were aligned in AliView with reference CAP277 Illumina sequences generated from plasma RNA obtained from Ms. Lynn Tyers (Division of Medical Virology, Faculty of Health Sciences, UCT) and analysed on MEGA X (Kumar *et al.*, 2018) using a Neighbour-Joining phylogeny. All 6 sequences were highly homologous with the CAP277 reference sequences (data not shown). Additionally, and surprisingly, none of the proviral sequences from any of the participants were found to be significantly hypermutated following screening using the Hypermut tool (LANL, 2014).

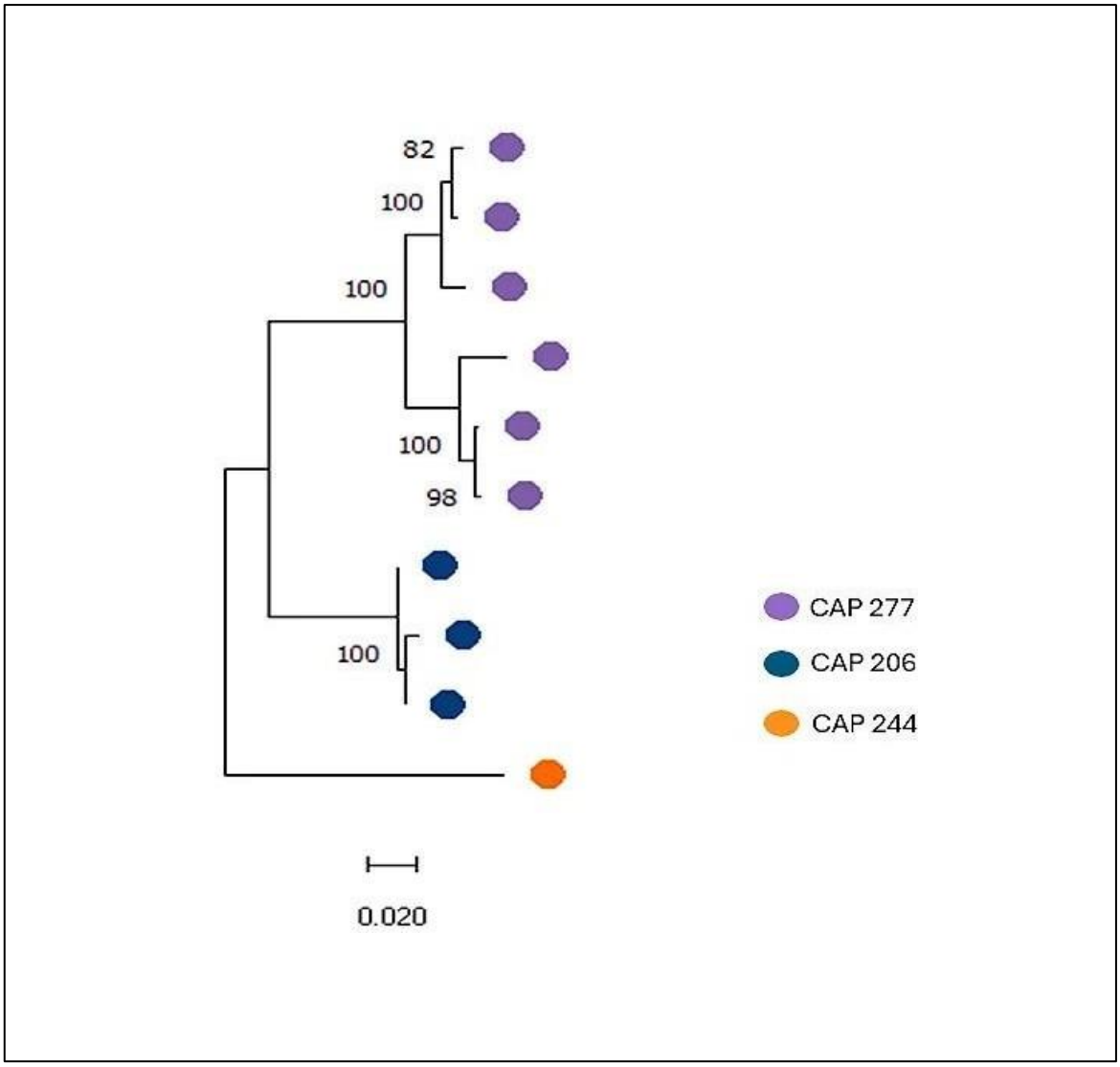

Figure 3.11. Neighbour-joining phylogeny of *env* V1V2 sequences generated in MEGA X illustrating clustering patterns of CAPRISA sample sequences and bootstrap values with confidence above or equal to 75.

DNA sequences from each participant with multiple variants sequenced (CAP277 and CAP206) clustered together on a phylogenetic tree with a bootstrap value of 100 (**Figure 3.11**). CAP244 had only one sequence but grouped separately.

3.5 *env V1V2* amplification and sequencing of HIV naïve samples spiked with 8E5 and SUP-T1 cells

To evaluate sensitivity of the Illumina Miseq NGS platform at detecting and sequencing low levels of HIV-1 DNA in breastmilk, HIV naïve breastmilk samples were spiked with 8E5 cells at concentrations ranging from 20 000 to 2 8E5 genome copies/million cells, following a 10-fold serial dilution from breastmilk samples BM8 A-E and supplemented with SUP-T1 cells to maintain the total number of cells at 5 million per sample (section 2.4.3, Methods). Genomic DNA from these samples were PCR amplified (section 3.4). For each sample, 1000 ng of DNA was amplified, followed by NGS sequencing.

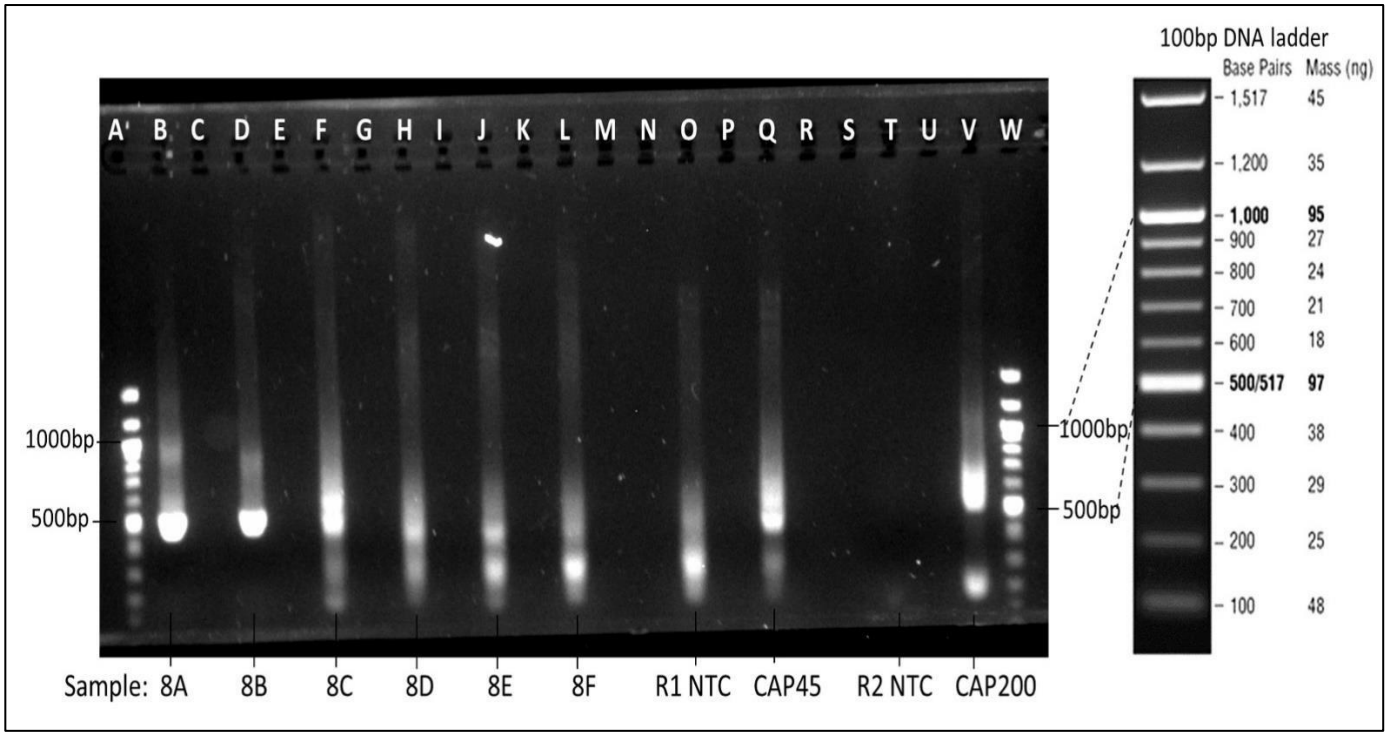

Figure 3.12. 1% Agarose gel image of multiplexed PCR of spiked breastmilk samples BM8 A-F following second round/indexing PCR. Lanes B, D, F, H, J and L contained samples BM 8A, 8B, 8C, 8D, 8E and 8F respectively. Lane O contained the indexed product from inputting 2 µl of the first round (R1) NTC to control for bands generated due to the presence of residual first round PCR primers, while Lane T contained a PCR second round (R2) NTC. Lanes Q and V contained indexed product from inputting 0.5 µl of CAP45 IMC at 2 pg/µl and CAP 200 plasma RNA first round PCR product as PCR R1 and PCR second round (R2) positive controls respectively. Lanes A and W contain a 100 bp DNA ladder (Quick-Load® 100 bp DNA Ladder, New England BioLabs Inc.) with GelGreen™ dye, also illustrated on the right.

Samples BM 8A, 8B and 8C in lanes B, D and F, respectively (**Figure 3.12**) produced the strongest DNA amplification within the expected band range. These samples corresponded to 8E5 spike-in concentrations of 20 000, 2000 and 200 proviral copies/million cells respectively. Sample BM8C produced a double band much like was seen for optimisation of CAPRISA sample CAP206. From samples BM 8D to 8F, the bands produced seem to decrease in size, with a double band in 8E, likely corresponding to bands in the non-spiked (negative control) and NTC reactions. As observed in several control wells from previous experiments, bands below 100 bp (miniPCR bio, no date) were likely primer dimers and bands below 500 bp but

higher than 100 bp were likely a result of interaction between first and second round PCR primers, which share common stretches of nucleotides, as seen after adding R1 NTC product into a second round PCR reaction (lane O) and the non-spiked sample BM8F (lane L). The absence of bands between 500 bp and 1000 bp indicates that lanes O and T PCR R1 and R2 NTCs are free of contamination while lanes Q and V PCR R1 and R2 positive controls successfully produced bands within the expected range.

The target 8E5 HIV-1/LAV sequence was successfully sequenced for samples BM 8A, 8B, 8C and 8D only. Sequences passing a quality control score of 25 (Illumina, 2011) were analysed similarly to section 3.4.3 and the results shown in **Table 3.4**.

Table 3.4. 8E5 spiked breastmilk sample sequences outputs

Sample (BM)	DNA input/ng	Number of assembled (Q score >25) sequences	Number of HIV-1/LAV sequences collapsed	Number of HIV-1/LAV consensus sequences after applying 1% frequency cutoff
8A	1000	327854	86594	1
8B	1000	369430	93273	1
8C	1000	186933	50535	1
8D	1000	62115	2986	1

Following sequence analysis, the raw read depth ranged from 410 472 to 1 433 263 and the mean quality score of forward and reverse reads ranged from 27.83 to 36.06 and 26.46 to 32.14 respectively. After merging the forward and reverse reads, the number of sequences ranged from 62 115 to 369 430. Identical sequences were collapsed using the dereplicate function in VSEARCH and sequences present at less than 1% frequency post dereplication were excluded. As expected, due to the clonal nature of the 8E5 cell line, all sequences dereplicated into a single variant.

3.6 Quantifying and sequencing HIV in breastmilk samples from the DolPHIN-2 study

To measure and sequence HIV from the breastmilk compartment in the presence of ART, five breastmilk samples were obtained from the DolPHIN-2 study, which investigated safety and efficacy of Dolutegravir integrase inhibitor-based ART initiated in late pregnancy (Kintu *et al.*, 2020). DolPHIN-2 breastmilk cell pellets obtained from 2 ml of breastmilk samples were provided in duplicate by Mr Nai-Chung Hu (Division of Public Health, Faculty of Health Sciences, UCT). The cell count and viability of the breastmilk samples were unknown.

DolPHIN-2 breastmilk cell samples were selected based on known, corresponding blood plasma viral loads. We selected samples with blood plasma viral loads ranging from 4log $_{10}$ RNA copies/ml representing active viraemia to 1log$_{10}$ copies/ml, representing suppressed viral load (**Table 3.5**). DNA was extracted, precipitated and concentrations measured using the high sensitivity Qubit kit as previously described (sections 2.9.1 and 2.9.2, Methods). DNA yields are shown in **Table 3.5** below.

Table 3.5. Qubit concentrations of DolPHIN-2 gDNA samples.

DolPHIN-2 Sample	Blood viral load (copies/ml)	Qubit concentration (ng/µl)
25008	20	95.8
25012	3642	17.1
25028	29680	46.2
25054	331	43.0
25105	39	110

Viral loads ranged from 20 copies/ml to 29680 copies/ml for these samples, and the extracted DNA sample concentrations ranged from 17.1 ng/µl to 110 ng/µl, with a median of 46.20 ng/µl and average of 62.42 ng/µl.

3.6.1 Droplet digital PCR of DolPHIN-2 breastmilk samples

In order to assess the sensitivity of the ddPCR platform at detecting and quantifying HIV DNA present in breastmilk samples on ART, *gag, pol* and *rpp30* genes were measured from DolPHIN-2 samples where subtype C *gag* and *pol* primers and probes were multiplexed. These primers previously worked consistently in multiplex on subtype C samples within the HIV Diversity Group. To quantify *gag* and *pol* genes*, t*he maximum possible amount of gDNA was used from each sample in duplicate (737.7 ng, 355.7 ng and 331.1 ng for samples 25008, 25054 and 25105 respectively), however not exceeding 1000 ng (to prevent inhibition of droplet formation and oversaturation of amplitude signals (Fish *et al.*, 2022). For *rpp30* quantification, 100 ng of each sample was used. 8E5 positive controls and water NTCs were used in both cases, as illustrated in **Figure A6**, Appendices. Samples 25012 and 25028 had to be excluded due to insufficient volume.

While 1 - 2 positive droplets were present in one or more replicate wells for these samples, according to the NTC Poisson confidence maximum value of 0.25, none of the *gag* and *pol* concentrations exceeded background levels, hence proviral copies (reservoir size) could not be calculated. The *pol*, *gag* and *rpp30* concentrations of the DolPHIN-2 samples ranged from an average of 0.043 copies/µl to 0.085 copies/µl, 0.034 copies/µl to 0.085 copies/µl and 45.8 copies/µl to 157 copies/µl, respectively. The 8E5 positive controls produced an average of 772 *pol* copies/µl, 803 *gag* copies/µl and 1039 *rpp30* copies/µl. Thus overall, while very low-level HIV DNA amplification signal was detectable from all the DolPHIN-2 samples, none were above background for the experiment. Furthermore, *rpp30* copy numbers were unexpectedly low for the DolPHIN-2 samples relative to the positive control, despite the same DNA ng input.

Moreover, 2-LTR circle amplification was not included since the latter is present at very low levels compared to proviral copies (Henrich *et al.*, 2012) and proviral copies detected were too low to be significant.

3.6.2 *env* V1V2 and *int* gene amplification and sequencing of DolPHIN-2 genomic DNA study samples

To evaluate the sensitivity of Illumina MiSeq at sequencing low-level HIV-1 proviral DNA from breastmilk cells of women on ART and to investigate the dynamics of the HIV reservoir in the

breastmilk compartment, DolPHIN-2 samples 25008, 25054, 25096 and 25105 were used. These samples had maximum possible extracted DNA inputs of 1000 ng, 22.2 ng, 182 ng and 934.4 ng, respectively. They were amplified (*env* and *int* genes in multiplex) using the optimised Illumina MiSeq method described in section 3.4. DolPHIN-2 sample 25096 with viral load of 20 copies/ml and Qubit concentration of 10.4 ng/µl was included in this experiment, as shown in **Figure 3.14**.

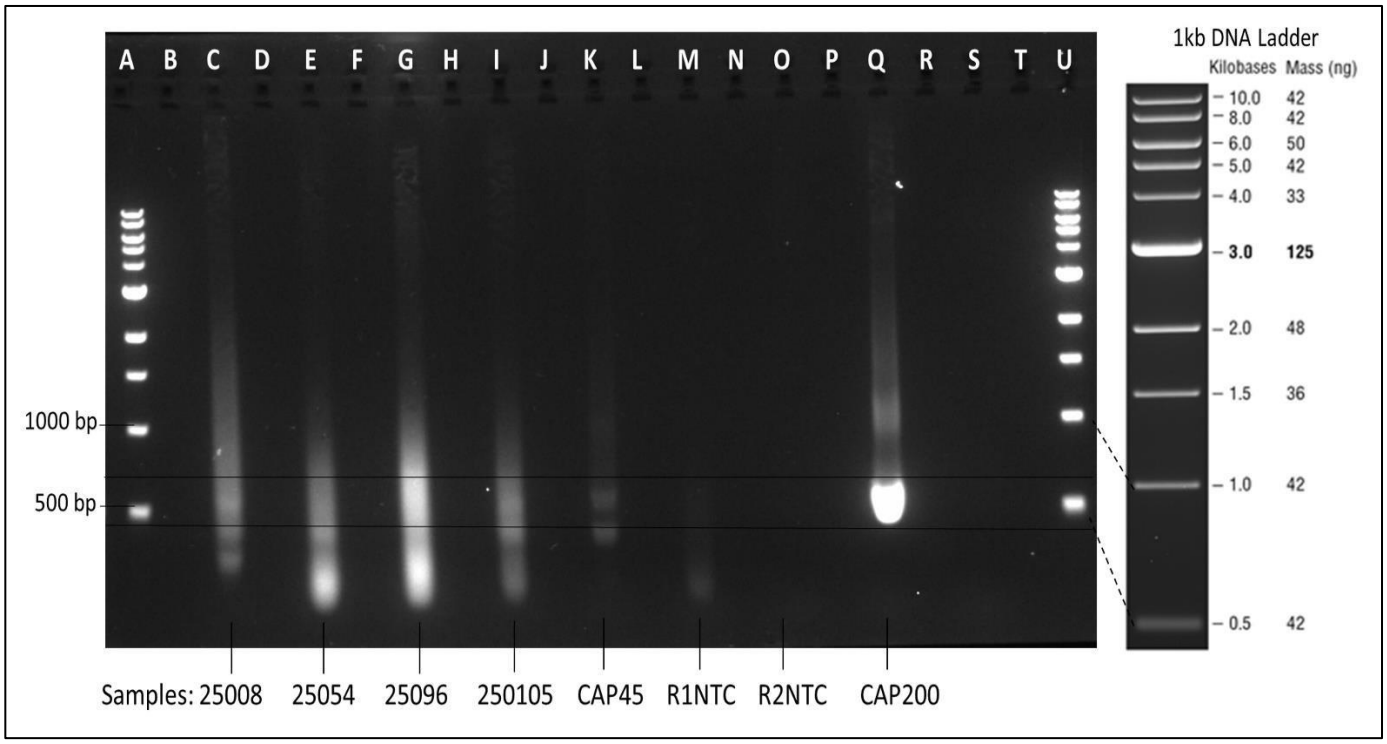

Figure 3.13. 1**%** Agarose gel image of multiplexed PCR of DolPHIN-2 samples following second round/indexing PCR with GelGreenTM dye**.** Lanes C, E, G and I contain DolPHIN-2 samples 25008, 25054, 25096 and 25105 respectively. Lane M contained the indexed product from inputting 2 µl of the first round (R1) non template control (NTC) to control for bands generated due to the presence of residual first round PCR primers, while Lane O contained a PCR second round (R2) NTC. Lanes K and Q contained indexed product from inputting 0.5 µl of CAP45 IMC at 2 pg/µl and CAP 200 plasma RNA first round PCR product as PCR R1 and PCR second round (R2) positive controls respectively. Lanes A and U contain a 1 kb DNA ladder (Quick-Load® 1kb DNA Ladder, New England BioLabs Inc.) with GelRed™ dye (Biotium), as shown on the right.

Regarding the DolPHIN-2 samples, while smearing was observed for all participant samples and potential faint DNA bands were observed for samples 25008 and 25015 at approximately 500 bp after PCR, no clear DNA bands were visible for any participant samples (**Figure 3.13**). Nonetheless, the band produced by CAP200 positive control around 500 bp was used as reference to excise DNA from an agarose gel for the DolPHIN-2 samples, as illustrated by the area between the black lines in the above figure. The CAP45 positive control produced a double band which could be due to both *env* and *int* PCR products being present. The NTCs were clear and the lower bands observed are primer dimers.

We attempted to sequence HIV-1 *env V1V2* and *int* regions to assess the Illumina MiSeq sensitivity and to investigate Dolutegravir drug resistance mutations in proviral DNA respectively. None of the samples produced the expected HIV sequences. These results are congruent to the ddPCR analysis of DolPHIN-2 samples where no HIV copies above background levels could be detected.

Chapter 4: Discussion

In the era of maternal ART, breastfeeding remains an important cause of infant HIV infection and accounts for 30 - 50% of all MTCT of HIV (Prendergast *et al.*, 2019). This is, in part, due to the presence of latent viral reservoirs in the circulating cells and mammary tissue despite the use of ART. Proviral HIV DNA in these reservoirs remains unaffected by ART and can persist for many years. Accurate and reliable methods are needed to characterize the breastmilk viral reservoir to inform transmission risk in the presence of ART.

We assessed the sensitivity of two platforms, the droplet digital PCR (ddPCR) and Illumina Miseq NGS platforms, to respectively quantify and sequence low-level HIV DNA in breastmilk. The use of these platforms in this context are novel, to our knowledge. This was performed by spiking HIV- naive breastmilk samples with 8E5 cells (each, harbouring a single copy of a defective HIV genome per cell) at a range of concentrations. ddPCR was able to detect intracellular HIV DNA ranging from spike-in concentrations of 20 000 - 20 8E5 genome copies/million cells with a limit of detection of 7 *gag* copies in a single reaction. However, this was dependent on total number of cells spiked. All viral DNA extracts derived from breastmilk samples that were quantifiable by ddPCR were also successfully sequenced by Illumina Miseq, thus confirming the presence of 8E5 genome sequences in the spiked-in samples. These methods were also applied to breastmilk samples from the Dolutegravir in Pregnant mothers with HIV infection and their Neonates (DolPHIN-2) study where HIV-positive breastfeeding women were administered DTG-based ART. The ddPCR results revealed intracellular HIV DNA copies below background levels, which were insufficient for Illumina MiSeq sequencing.

We determined the ideal spiking conditions for optimal DNA recovery for ddPCR and Illumina sequencing. As reported by Strain *et al.* for experiments with PBMCs, an optimal amount of 1 µg of gDNA input was required for each HIV target reaction wherever possible, to maximise sensitivity for low-level HIV detection, while input above 1 µg may result in PCR inhibition (Strain *et al.*, 2013).

One of the biggest challenges in our study was the recovery of sufficient gDNA for optimal ddPCR amplification of HIV targets. When we spiked breastmilk samples in a background of 1 million cells total prior to extraction, 1 µg of gDNA could be recovered and inputted per ddPCR

reaction only for samples spiked with the two highest 8E5 dilution concentrations (10 000 and 1000 8E5 cells/million cells respectively). In contrast, when spiking breastmilk in a background of 5 million cells total, all spiked samples provided 1 µg of gDNA per reaction for ddPCR quantification. This produced higher extracted gDNA concentrations overall, with an average fold increase of 9.4 as compared to the previous spike-in experiment. Sufficient template DNA could not be obtained from samples for estimating viral spike-in concentrations in the absence of background cells. Pasteurization of breastmilk most likely resulted in cellular death and thus poor cellular DNA recovery in this instance. In future, spiking with background cells is therefore needed for sufficient DNA recovery for these types of experiments. We found 5 million cells to be optimal as used by Strain *et al.* (2013) to quantify HIV DNA from PBMCs.

Ghosh *et al.* performed a similar experiment where they spiked HIV-negative cell-free breastmilk with 8E5 cells following a $\log_{10}$ serial dilution from inputs ranging from 1 to 1 million ($6\log_{10}$) cells and added PBMCs or breastmilk cells to maintain a constant number of 1 million cells per ml of breastmilk. Using real-time PCR (TaqMan), they detected one HIV copy in 2 out of 10 (20%) assays and 10 HIV DNA copies ($1\log_{10}$) in 8 out of 10 assays (80% sensitivity) but did not report total DNA input per reaction making direct comparison of our studies challenging. In contrast, we were only able to detect HIV DNA down to $1.3\log_{10}$ when extracting from a total of 5 million cells, albeit in two out of two replicate reactions. Unlike our experiment, Ghosh *et al.* (2003) applied real-time PCR instead of ddPCR.

Additionally, from the positive control for this experiment, we observed an average of 0.8 HIV copies/ 8E5 cell as opposed to the expected 1 8E5 HIV copies/ 8E5 cell (ATCC, 2024; Désiré *et al.*, 2001). Similar results were observed by Busby *et al.* They cultured 8E5 cell lines obtained from different sources for quantification by qPCR and ddPCR. The study reported a range of 0.02 to 0.8 HIV DNA copies/8E5 cell, indicating that different batches of 8E5 cells may contain different amounts of HIV DNA per cell. Passaging of a fresh culture of 8E5 cells from the ATCC revealed that with each successive passage, HIV DNA copies decreased relative to a human reference gene, resulting in a reduction from about 0.8 to 0.6 HIV DNA copies per cell over four passages (Busby *et al.*, 2017). Similarly, we observed 8E5 genome copies ranging from 0.4 to 0.8 HIV copies/cell across our experiments. The HIV DNA loss mechanism is not clear, but it is known that the site of proviral insertion in 8E5 cells (13q14-q21) is fragile (Deichmann *et al.*, 1997) and could possibly contribute to 8E5 proviral loss through genomic instability

(Busby *et al.*, 2017). For future spiking studies, J-lats may be a better alternative to 8E5 cells since the former have been used in several HIV latency (reversal) studies (Burnett *et al.*, 2010; Mbonye and Karn, 2017; Spina *et al.*, 2013) and none of them have reported genome instability.

Due to this observed instability of viral genome copy numbers in the 8E5 cell line, we conservatively spiked breastmilk samples in our final experiment with concentrations of 50 000 - 5 8E5 cells/million following a 10-fold serial dilution. The 8E5 positive control from this experiment revealed a measured average of 0.4 *gag* copies/cell. Therefore, it could be extrapolated that only 0.4x the number of HIV copies were in fact spiked-in, which is 20 000 – 2 8E5 HIV genomes/million cells.

Furthermore, increasing the total number of cells resulted in detectable HIV copies from samples spiked with 50 to 50 000 8E5 cells/million cells (BM8 A-E), which corresponded to intracellular DNA copy concentrations ranging from 44.29 to 9239.97 HIV copies/million cells where 44.29 HIV copies/million cells corresponded to 7 *gag* copies in a single ddPCR reaction. Defined as the template concentration below which detection will usually not occur at levels distinguishable from background noise, Strain *et al.* reported a limit of detection between 3-4 HIV copies/million cells depending on the number of droplets formed (Strain *et al.*, 2013). Our sensitivity for low copy number detection was therefore lower than reported by Strain *et al*. and Ghosh *et al*. (2003) in both blood and breastmilk, respectively. However, we did not plan to explore sensitivity over a large number of replicates in this study, but this may be investigated in future work.

Intracellular HIV concentration from the spiked breastmilk samples followed the dilution series until deviation at the 20 8E5 HIV genomes/million cells (50 8E5 cells/million) spike-in, possibly indicating a loss of accuracy and sensitivity at this point. Moreover, from samples BM8 A-C, the measured spike-in concentrations in 8E5 copies/ million cells were higher than the intracellular DNA concentration in average *gag* copies/million cells by 2-fold. For instance, while 20 000 8E5 genome copies/million cells were spiked in, 9239.97 HIV copies/million cells were detected. This discrepancy could be due to differences between the calculated and actual number of 8E5 cells spiked-in possibly due to instrumental error from the Countess™ or cell loss during dilution, leading to less cells being spiked-in. Indeed, the automated cell counter may overestimate or underestimate cell count (Conduct Science, 2024) although with

less than 5% count-to-count variability (Invitrogen, 2015). Additionally, the instrument is less efficient at detecting atypical cells. Hence, results may vary due to these factors (Conduct Science, 2024).

Another possible explanation could be inaccuracies with the NanoDrop spectrophotometer. In fact, we also observed fluctuations in *rpp30* concentrations across spiked breastmilk samples BM6 A-F while similar concentrations were expected since all samples contained 1 million cells supplemented with PBMCs, and *rpp30* is a highly conserved housekeeping gene which is widely expressed in human tissues (Wu *et al.*, 2022). Such variation may also have originated from inaccuracies during quantification of HIV concentration by the NanoDrop light spectrophotometer. The NanoDrop light spectrophotometer exhibits a lack of sensitivity at quantifying samples with low concentration, particularly below 10 ng/µl (AAT Bioquest, 2022) despite a reported limit of detection of 2 ng/µl (Thermo Fisher Scientific, 2016). Furthermore, sample DNA concentrations may have been overestimated since DNA and RNA both absorb at 260nm (AAT Bioquest, 2022).

Considering the issues observed with the use of the NanoDrop, Qubit may be a better alternative at quantifying DNA. Indeed, we observed different DNA concentration readings when using the NanoDrop and Qubit on the same samples whereby NanoDrop readings were usually higher than Qubit readings (data not shown), a phenomenon also observed by Masago *et al.* Nanodrop measures overall DNA present in a sample while Qubit only measures double stranded DNA specifically (Masago *et al.*, 2021). Moreover, Qubit demonstrates more reliable and accurate data than the NanoDrop with measurement errors around 1% and 5% respectively when quantifying 10 ng/µl of DNA samples (Masago *et al.*, 2021; Sah *et al.*, 2013). An interesting recommendation made by Simbolo *et al.* proposes that for adequate DNA quantification for NGS, the sample quality can first be assessed by NanoDrop and DNA measured using Qubit (Simbolo *et al.*, 2013). Hence, a combination of both instruments may produce optimal results.

Nonetheless, the NanoDrop 2000 spectrophotometer remains a valuable tool for quick and convenient nucleic acid and protein quantification. Furthermore, ddPCR methodology is based on a study carried out by Strain *et al.* (2013) who used this instrument for the quantification of DNA.

To evaluate sensitivity of the Illumina Miseq NGS platform at detecting and sequencing low levels of HIV-1 DNA in breastmilk samples, the envelope gene V1-V2 region was amplified. The V1-V2 region, located in gp120 of the envelope apex, harbours several glycans and is highly sequence diverse (Pan *et al.*, 2015). Sequencing this region enables us to distinguish between viral variants in a swarm of quasispecies.

We were successful at amplifying and sequencing HIV-1 from gDNA derived from PBMC of women suppressed on ART (from the CAPRISA study cohort) with a depth of 59 105 – 211 749 sequences recovered from three women. The previously determined reservoir sizes for these women ranged from 82 - 905 *gag* copies/million cells and 92 – 1309 *pol* copies/million, respectively, demonstrating the sensitivity of the Illumina MiSeq method for characterising proviral populations in clinical samples from PLHW on ART. No hypermutation was found in these sequences despite reports that 93% of HIV proviruses are defective with large deletions and hypermutation (Bruner *et al.*, 2016).

Regarding 8E5 cells spiked in breastmilk samples, we successfully amplified and sequenced HIV-1 at a range of 44.3 – 9240.0 *gag* copies/million cells and the sequences generated corresponded to the 8E5 genome sequence. This once again demonstrated sensitivity of the Illumina MiSeq platform at sequencing HIV DNA from breastmilk.

However, when applying this method to gDNA obtained from breastmilk cell pellets of breastfeeding women on Dolutegravir-based ART (from the DolPHIN-2 study) with blood viral loads ranging from 20 - 331 copies/ml (breastmilk viral loads were unknown), no amplification was observed, possibly due to low HIV-infected cell numbers in the breastmilk.

Regarding the CAPRISA samples, our recorded raw read depth (59 105 – 211 749 reads) was lower than optimal (300 000 - 400 000 reads) (Abrahams, 2024, personal communication), suggesting that further optimisation may be needed. Furthermore, we could not apply the Primer ID (PID) approach for distinguishing viral DNA templates and controlling for PCR bias (Jabara *et al.,* 2011) as previously used for Illumina MiSeq amplification from RNA in this cohort (Abrahams *et al*., 2019; Joseph *et al.,* 2024) as PIDs are usually introduced through cDNA synthesis from RNA templates. Instead, we collapsed sequences to represent unique viral variants present at frequencies greater than 1%, a common practice for Illumina MiSeq data (Howison, Coetzer and Kantor, 2018). This resulted in very few unique variants for each

CAPRISA participant sample, and we cannot be certain whether this is due to high clonality of the proviral population, which we could not differentiate from PCR bias, or simply the rarity of proviral copies in these samples.

For comparison, a pre-ART sample from the same cohort with very low blood plasma RNA viral load of 483 copies/ml (approximately 193 RNA copies used for PCR) produced 28 consensus sequences obtained after collapsing on PID and 8 unique sequences following collapsing on 100% identity of variants (Tyers, 2024, personal communication). In comparison, we obtained 1, 3 and 6 unique variants from CAPRISA samples (on ART) corresponding to medium, high and low HIV-1 reservoir sizes (484, 1309 and 92 *pol* copies/million cells respectively). Although not a direct comparison, few sequences can be expected from samples with low HIV-1 levels. Moreover, we sequenced the envelope region which is mostly defective in proviral populations in the reservoir (Wei *et al.*, 2020), potentially resulting in under-representation of variants from the reservoir.

Interestingly, we recorded *rpp30* concentrations ranging from only 45.8 copies/µl to 157 copies/µl from the DolPHIN-2 samples which were unexpectedly low in comparison to 1039 *rpp30* copies/µl obtained from the 8E5 positive control with same DNA input (100 ng). This disparity may have significantly reduced our sensitivity.

Low *rpp30* copies in DolPHIN-2 samples may indicate low cell number from these samples. Nonetheless, some DolPHIN-2 samples had relatively high DNA concentrations, up to 110 ng/µl which may have mainly been background DNA since HIV target was very low. This could possibly be attributed to a diverse microbiota in breastmilk. Indeed, breastmilk is not sterile and contains more than 800 bacterial species (Togo *et al*., 2019) including beneficial, commensal, and potentially probiotic bacteria (Lyons *et al.*, 2020) which colonise the infant's gastrointestinal tract for healthy gut development (Bode *et al.*, 2014; Notarbartolo *et al.*, 2022).

Since we know that the 2 ml of expressed breastmilk from the DolPHIN-2 study was centrifuged and cells were pelleted, we could speculate that this volume of breastmilk may have contained insufficient cells to detect HIV copies above background levels in individuals on ART as we observed from our spiking experiments where a total of 5 million cells spiked at $1.3\log_{10}$ and above HIV copies/million cells resulted in detectable HIV.

Lack of HIV detection recorded may also be due to too low blood viral load and possibly even lower breastmilk viral load. Indeed, among untreated HIV-positive mothers, it has been reported that the average and median breastmilk viral loads were lower than that of maternal blood plasma by $2\log_{10}$ and 1.68-fold difference respectively (Kuhn *et al.*, 2013; Rousseau *et al.*, 2003). Also, the local host environment in different tissues may significantly affect the size, composition and distribution of the reservoir during ART. Thus, studies of peripheral blood may not reflect what happens in tissues, including the breast (Svicher *et al.*, 2014).

Another possible explanation for lack of HIV detection could be high efficacy of DTG-based ART. DTG has proven to be safe and effective at reducing viral load in HIV-infected pregnant women to undetectable levels twice as fast as conventional Efavirenz based regimens (Dickinson *et al.*, 2021; Kintu *et al.*, 2020; Malaba *et al.*, 2022; Waitt *et al.*, 2019).

Overall, we cannot rule out that breastmilk processing may have affected HIV DNA recovery, quantification and sequencing. For instance, for the spiking experiments, HIV negative breastmilk samples were pasteurised using the Holder pasteurisation method and stored in a 4˚C fridge for about a week until used or discarded. For the first spiking experiment with 8E5 cells only, cell viability of HIV-negative breastmilk samples was measured and ranged from 7% to 42%. Furthermore, breastmilk cells were pelleted and stored at –80˚C (for an extended period for DolPHIN-2 samples). Pasteurisation and potentially storage conditions may have reduced cell viability of our samples. Additionally, removal of the fat layer in breastmilk was difficult, particularly without any known standardized guidelines available. We found that storage of spiked samples overnight at 4˚C and pelleting of cells at 4˚C compacted the fat layer making removal somewhat easier.

There are varying methods of breastmilk storage and handling reported in the literature. For example, in a SARS-CoV-2 study carried out by Groß *et al.,* breastmilk samples were stored at 4˚C, then at -20˚C or at -20˚C directly or analysed on the same day of milk expression without storage at -20˚C. They prepared skim milk by centrifugation for 10 minutes at 1000 × g and analysed the lower aqueous phase (Groß *et al.*, 2020). This is similar to our approach with a much lower centrifugation speed. Bäuerl *et al.* also used expressed milk, which they stored at −20°C and -80°C for SARS-CoV-2 RNA detection and thawed and centrifuged milk samples at 14 000 rpm for 10 minutes at 4˚C (like we did but) with centrifugation repeated twice to remove all fat and cells and frozen at -80°C for further analysis (Bäuerl *et al.*, 2022). They were

able to successfully demonstrate the safety of breastfeeding from mothers with SARS-CoV-2 using this method although Groß *et al.* detected SARS-CoV-2 RNA in their breastmilk samples and reported better recovery rates with skim milk as compared to whole milk for viral detection by qPCR (Groß *et al.*, 2020).

In the light of the COVID-19 pandemic, McGuire *et al.* have proposed best practices for breastmilk collection and storage. For milk partitioning and storage, expressed milk is recommended to be refrigerated immediately. For the isolation of infectious virus, the sample might need to be analysed immediately or snap frozen. It is also recommended to aliquot milk in sterile storage containers to prevent repeated freeze-thaw cycles followed by freezing preferably at -80˚C or at -20˚C (McGuire *et al.*, 2021).

Additionally, breastmilk cell pellets contain milk lipid globule membranes, proteins and glycoproteins such as lactoferrin and immunoglobulin G, but mainly fat-laden macrophages which may inhibit PCR amplification (Ghosh *et al.*, 2003). Reporting on the matter, Ghosh *et al.* and Shepard *et al.* reported partial inhibition in one-third and three of their five breastmilk samples respectively but more than 90% of the samples had sufficient recovery to obtain valid results. Using real-time PCR, Ghosh *et al.* therefore demonstrated that despite the presence of breastmilk components that may inhibit PCR quantification, HIV can be accurately detected in 8E5-spiked breast milk samples and samples from HIV-infected women.

Similarly, despite challenges with the fat layer, we do not believe that this significantly impacted our sensitivity for detection of HIV as similarly to Ghosh *et al.*, HIV was successfully quantified in breastmilk samples where sufficient DNA was available for ddPCR detection from our spiking experiments.

Prior to breastmilk spike-in experiments, we evaluated the ability to detect 8E5 cell line (subtype B) HIV DNA using ddPCR HIV subtype C *gag* and *pol* primers and probes from existing in-house assays. Alignment with the HIV-1/LAV/LAI genome reference revealed good compatibility between the reference and existing HIV subtype C *gag* primers and probes but not with *pol* primers and probes which were found to contain several mismatches, including at the 3’ end of primers.

Sequence mismatches in primer targets may lead to under-quantification or non-specificity, resulting significant ramifications for viral quantification studies. For instance, Haqshenas *et*

al. reported off-target amplification and under-detection of target due to nucleotide mismatches at primer binding sites during ddPCR (Haqshenas *et al.*, 2023) while Whiley and Sloots reported delayed amplification and viral load underestimation with qPCR (Whiley and Sloots, 2005). Additionally, it has been shown that mismatches towards the 3' end of a primer reduces target amplification efficiency much more than mismatches towards the 5' end (Ghedira *et al.*, 2009; Whiley and Sloots, 2005; Ye *et al.*, 2012) since DNA polymerase adds nucleotides from the 3' end during replication (Huang *et al.*, 2024).

Therefore, for our experiments, previously published subtype B *pol* primers and probes (Strain *et al.*, 2013) were used instead as those demonstrated 100% match to the 8E5 cell line reference sequence. While the subtype C *gag* reverse primer demonstrated 100% compatibility, two mismatches were found in the *gag* forward primer at the 5' end and one mismatch in the middle of the *gag* probe.

Despite previous studies finding that a mismatch in the centre of a probe result in destabilized probe-template annealing (Süß *et al.*, 2009; Winton *et al.,* 2002), amplification can still occur (Ye *et al.*, 2012), depending on their location, hence existing *gag* primers and probes were used. Moreover, in contrast to qPCR, ddPCR is less affected by mismatches due to its end-point detection technique (Kojabad *et al.*, 2021; Persson *et al.*, 2019). Reduced amplification efficacy in ddPCR is reflected in the production of droplets with lower fluorescence intensity (Persson *et al.*, 2019). Nonetheless, quantification is generally unaffected for positive droplets above the ddPCR amplitude threshold.

Indeed, in this project we observed amplification of *gag* from 8E5 cell gDNA despite the single mismatch identified. Furthermore, when multiplexing HIV subtype B *pol* primers and probes and well as subtype C *gag* primers and probes for HIV quantification in 8E5 cells following a dilution series, for 3 out of 4 dilutions, the *pol* and *gag* concentration (copies/µl) recorded were highly similar at their corresponding 8E5 dilutions (data not shown). Thus, we can be confident that amplification at a reduced level with this primer/probe set was unlikely.

Finally, with regards our model used for HIV spike-in experiments, while multiplexing of HIV amplification for 8E5 genomic DNA was desirable, we encountered what was likely probe cross-reactivity during ddPCR optimization. When multiplexing *gag* and *pol* probes and primers to quantify 8E5 DNA from 50 ng to 0.4 ng reaction inputs following a 5-fold dilution

series, double amplitude bands were observed for both positive and negative droplets. Furthermore, double positive clouds were observed when multiplexing *pol* and *rpp30* genes to quantify 8E5 cells from 100 ng to 0.16 ng DNA input following a 5-fold dilution series. In contrast, this was not observed when quantifying gDNA derived from women on ART infected with HIV subtype C (Sherazaan Dineo Ismail, 2022). The main difference between these cases other than HIV-1 subtype was target abundance. When quantifying from 8E5 cells, target levels were much higher and genomes are expected to be intact while the on-ART samples are highly likely to harbour low-level, predominantly defective HIV genomes, thus limiting primers/probes from binding to the target regions.

Therefore, 8E5 cells with high target concentration may promote probe cross-reactivity. In fact, high template concentrations can promote non-specific binding of primers and probes, increasing the chances of amplifying off-target sequences (Dube, Qin and Ramakrishnan, 2008). When a probe binds to a mismatched sequence, it may undergo cleavage, resulting in cross-reactivity (BIO-RAD, no date). Consequently, we performed singleplexed ddPCR reactions.

Ultimately, in spiked-in breastmilk samples, we detected low target DNA despite recording high gDNA concentration, thus most of the sample DNA was probably background DNA. A method that either isolates and enriches HIV susceptible cells (e.g. CD4+ T cells, particularly when these cells are rare in a particular compartment) before DNA extraction or enriches for HIV target would be ideal for HIV quantification and characterization in breastmilk.

For instance, Fluorescence-Activated Cell Sorting (FACS) or magnetic-activated cell sorting (MACS) can be used in combination with ddPCR to sort HIV-susceptible cells, such as CD4+ T cells and quantify HIV DNA respectively. These techniques use fluorescently labelled or magnetic bead bound antibodies to target specific cell surface antigens which get isolated and collected. Moreover, probe capture enrichment is another method that can be used to sequence samples containing low genomic copy numbers. After nucleic acid extraction, a library is prepared and custom-designed probes hybridise to the target within the template genome. Sequencing can be carried out by Illumina Miseq and Next Seq platforms which are commonly used in target enrichment studies (Munyuza, Ji and Lee, 2022). These methods have been used successfully in a wide array of pathogens, including HIV. Alternatively, larger breastmilk volumes could be used to obtain more breastmilk cells since we noticed that

pelleting cells from 2 ml of DolPHIN-2 milk samples was insufficient. Nonetheless, working with large breastmilk volumes is not practical in our settings due to volume limitations with the instruments used.

In conclusion, we demonstrate proof of concept that HIV can be detected and quantified by the ddPCR platform from HIV-negative breastmilk samples spiked with cells harbouring HIV genomes with a sensitivity of 7 HIV copies from 50 8E5 spiked-in cells/million (or 20 HIV copies/million cells). The 8E5 HIV DNA variant was successfully sequenced from the spiked-in breastmilk samples with the same sensitivity using the Illumina MiSeq platform.

Additionally, we showed that breastmilk samples obtained from HIV-positive mothers on Dolutegravir based ART did not harbour detectable HIV copies above background levels and Illumina MiSeq sequencing of HIV DNA was unsuccessful, thus illustrating potential low risk of HIV MTCT by breastfeeding in the presence of DTG. Nonetheless, only three samples from the DolPHIN-2 study were analysed and their corresponding blood viral loads were low. From our spiking experiments, we found that a total of 1 million cells was insufficient for DNA quantification and that ddPCR detection of HIV improved greatly with a total number of cells increased to 5 million. The number of cells in the DolPHIN-2 samples is unknown but can be assumed to be low based on lower-than-expected *rpp30* copies obtained from these samples.

Processing breastmilk samples for HIV quantification and sequencing is challenging since there is no established protocol or SOP and to our knowledge, and no one has yet managed to successfully quantify and sequence HIV DNA from breastmilk samples of HIV-positive mothers on ART with low viral loads, to our knowledge. Possible improvements for our study include increasing the number of clinical study samples analysed, quantifying breastmilk HIV DNA from participants with high blood viral load for proof of concept, increasing cell viability by using raw milk instead of pasteurised milk, use of J-Lats for spiking experiments, cryopreservation of expressed breastmilk samples in liquid nitrogen instead of at -80˚C as well as pelleting a larger number of cells by using a larger volume of breastmilk.

Overall, we propose the use of both ddPCR and Illumina Miseq as valuable platforms for the characterization of the latent HIV reservoir in breastmilk.

Appendices

Appendix 1: Breastmilk spiking experiments

1.1 Spiking breastmilk samples with 8E5 cells only

Table A1. Number of breastmilk cells needed from samples BM1, BM2 and BM3

Spiked sample ID	Number of BMCs needed*	Volume of BM1 (cell count $1.6x10^7$ BMCs/ml) needed (µl)	Volume of BM2 (cell count $3.5x10^6$ BMCs/ml) needed (µl)	Volume of BM3 (cell count $3.5x10^6$ BMCs/ml) needed (µl)
A	990000	61.9	283.0	283.0
B	999000	62.4	285.4	285.4
C	999900	62.5	285.7	285.7
D	999990	62.5	285.7	285.7
E	999999	62.5	285.7	285.7
F	$1x10^{\wedge 6}$	62.5	285.7	285.7

*Pipetting increments below 0.1 were not possible. Therefore, volumes from sample C to F were equivalent

1.2 Spiking breastmilk samples with 8E5 cells and PBMCs

Table A2. Number of PBMCs needed to spike breastmilk samples BM4 and BM6.

Spiked sample	Number of BMCs needed	Volume of PBMC-BC14 (cell count $2.8x10^6$ cells/ml) needed to spike BM4 (µl)	Volume of PBMC-BC15 (cell count $2.2x10^6$ PBMCs/ml) needed to spike BM6 (µl)

A	990000	353.6	450.0
B	999000	356.8	454.1
C	999900	357.1	454.5
D	999990	357.1	454.5
E	999999	357.1	454.5
F	1x10^6	357.1	454.5

*Pipetting increments below 0.1 were not possible. Therefore, volumes from sample C to F were equivalent

1.3 Spiking breastmilk samples with 8E5 cells and SUP-T1 cells

Table A3. Number of PBMCs needed to spike pooled breastmilk sample BM8.

Breastmilk sample	**Number of SUP-T1 cells needed**	**Volume of SUP-T1 (cell count $1.1x10^6$ PBMCs/ml) needed to spike BM8 (ml)**
A	4750000	4.32
B	4975000	4.52
C	4997500	4.54
D	4999750	4.55
E	4999975	4.55
F	5000000	4.55

*Pipetting increments below 0.1 were not possible. Therefore, volumes from sample D to F were equivalent

Appendix 2: ddPCR experiments

2.1 ddPCR primers and probes

Table A4. HIV-1 subtypes B and C ddPCR primers and probes

Primer/probe	Sequence	HXB2 genome positions
Pol Fwd 6	TCG GGT TTA TTA CAG AGA CAG CAG AGA	4898 – 4924
Pol Rev 4	AGC ICC TGC CAT CTG TTT TCC AT	5041 – 5063
Pol probe 3 FAM ZEN	/56-FAM/AAG GAC CAG/ZEN/CCA ARC TAC TCT GGA AAG GTG/3IABkFQ/	4936 – 4965
Gag Fwd	GTT GGA GGA CAT CAA GCA GCC ATG CA	1360 – 1385
Gag Rev	TTC CTG CTA TGT CAC TTC CCC T	1483 – 1504
Gag probe HEX ZEN	/5HEX/ACC ATC AAT/ZEN/GAR GAG GCT GCA GAA TGG GA/3IABkFQ/	1399 – 1427
RPP30 Fwd	GAT TTG GAC CTG CGA GCG	
RPP30 Rev	GCG GCT GTC TCC ACA AGT	
RPP30 probe HEX ZEN	/5HEX/CTG ACC TGA/ZEN/AGG CTC T/3IABkFQ/	
HIV pol forward (mf299)	GCA CTT TAA ATT TTC CCA TTA GTC CTA	2536 – 2562
HIV pol reverse (mf302)	CAA ATT TCT ACT AAT GCT TTT ATT TTT TC	2634 – 2662
HIV pol probe (mf348)	VIC-AAG CCA GGA ATG GAT GGC C-MGB	2586 – 2604

2.2 Optimization of ddPCR quantification of HIV-1 from 8E5 cells

2.2.1 Testing HIV-1 quantification from 8E5 cells using subtype B pol and subtype C gag primers and probes

Table A5. The components of *gag* and *pol* ddPCR mastermix for multiplexed reactions.

Component	Concentration	Volume per reaction (µl)	Volume for 8.3 reactions (µl)*

Supemix	2x	10.00	83.0
HIV pol forward (mf299)	50µM	0.18	1.5
HIV pol reverse (mf302)	50µM	0.18	1.5
HIV pol probe (mf348)	25µM	0.20	1.7
Gag_Fwd	50µM	0.18	1.5
Gag_Rev	50µM	0.18	1.5
Gag_probe HEX ZEN	25µM	0.20	1.7
Ban II enzyme	10000 U/ml	0.50	4.2
CutSmart buffer	10x	0.70	5.8
mg H_20		2.7	22.2
Template DNA		5	

*****Equivalent to one column of a 96-well plate with provision for pipetting error.

2.2.2 Testing HIV-1 quantification with rpp30 and pol primers and probes in multiplexed v/s singleplexed experiments.

Table A6. The *pol* ddPCR mastermix for singleplexed reactions.

Component	**Concentration**	**Volume per reaction (µl)**	**Volume for 8.3 reactions (µl)**
Supemix	2x	10	83
HIV pol forward (mf299)	10µM	0.9	7.5
HIV pol reverse (mf302)	10µM	0.9	7.5
HIV pol probe (mf348)	10µM	0.5	4.2
Ban II Enzyme	10000 U/ml	0.5	4.2
CutSmart buffer	10x	0.7	5.8
mg H_20		1.5	12.5
Template DNA		5	

Table A7. The *rpp30* ddPCR mastermix for singleplexed reactions.

Component	**Concentration**	**Volume per reaction (µl)**	**Volume for 8.3 reactions (µl)**
Supemix	2x	10	83
RPP30 Fwd	10µM	0.9	7.5
RPP30 Rev	10µM	0.9	7.5
RPP30 probe HEX ZEN	10µM	0.5	4.2
Ban II enzyme	10000 U/ml	0.5	4.2
CutSmart buffer	10x	0.7	5.8
mg H_20		1.5	12.5
Template DNA		5	

Table A8. The components of *rpp30* and *pol* ddPCR mastermix for multiplexed reactions.

Component	**Concentration**	**Volume per reaction (µl)**	**Volume for 8.3 reactions (µl)**
Supermix	2x	10.0	83.0
HIV pol forward (mf299)	50µM	0.18	1.5
HIV pol reverse (mf302)	50µM	0.18	1.5
HIV pol probe (mf348)	25µM	0.20	1.7
RPP30 Fwd	50µM	0.18	1.5
RPP30 Rev	50µM	0.18	1.5
RPP30 probe HEX ZEN	25µM	0.20	1.7
Ban II enzyme	10000U/ml	0.50	4.2
CutSmart buffer	10x	0.70	5.8
mg H_20		2.7	22.2
Template DNA		5	

2.3 Quantifying HIV-1 from DoLPHIN-2 breastmilk samples.

Table A9. *pol* ddPCR mastermix for quantifying breastmilk samples spiked with 8E5 cells and PBMCs.

Component	Concentration	Volume per reaction (µl)	Volume for 16.5 reactions (µl)
Supemix	2x	10.00	165.0
HIV pol forward (mf299)	50µM	0.18	3.0
HIV pol reverse (mf302)	50µM	0.18	3.0
HIV pol probe (mf348)	25µM	0.20	3.3
Ban II enzyme	10000 U/ml	0.50	8.3
CutSmart buffer	10X	0.70	11.6
Template DNA		8.2	

Table A10. *gag* ddPCR mastermix for quantifying breastmilk samples spiked with 8E5 cells and SUP-T1 cells.

Component	Concentration	Volume per reaction (µl)	Volume for 25 reactions (µl)
Supemix	2x	10.00	250.0
Gag Fwd	50µM	0.18	4.5
Gag Rev	50µM	0.18	4.5
Gag probe HEX ZEN	25µM	0.20	5.0
Ban II enzyme	10000 U/ml	0.50	12.5
CutSmart buffer	10x	0.70	17.5
Template DNA		8.2	

Table A11. *gag* and *pol* ddPCR mastermix for quantifying DolPHIN-2 study samples

Component	Concentration	Volume per reaction (µl)	Volume for 17 reactions (µl)
Supemix	2x	10.00	170.0
Pol Fwd 6	50µM	0.18	3.1
Pol Rev 4	50µM	0.18	3.1
Pol probe 3 FAM ZEN	25µM	0.20	3.4
Gag Fwd	50µM	0.18	3.1
Gag Rev	50µM	0.18	3.1
Gag probe HEX ZEN	25µM	0.20	3.4
Ban II enzyme	10000 U/ml	0.50	8.5
CutSmart buffer	10x	0.70	11.9
Template DNA		7.7	

The mastermix for *rpp30* remained as illustrated in **Table A7**.

Appendix 3: Ilumina MiSeq NGS amplifications

3.1 Illumina MiSeq PCR primer sequences

Table A12. PCR primer sequences for the *env* V1V2 and *int* regions as well as universal and indexed adapters.

Primer ID	Sequence*	HXB2 genome positions
V1F_UCT	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACA GNNNNATGGGATCAAAGCCTAAARCCATGTGTA	6515 - 6584
V2R_PID_UCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNN NNNNNNNNCAGTCTTAATTCCATGTGTACATTGTACT GTRCT	6951 - 7021

C9dn_UP_UCT	GCCTCCCTCGCGCCATCAGAGATGTGTATAAGAGACA GNNNNGGCTACATAGAAGCA GAGGT	4473 - 4492
C9dn_PID_UCT	GTGACTGGAGTTCAGACGTGTGCTCTTCCGATCTNNN NNNNNNNNCAGTAATCTTCATCCTGTCTACCTGCCAC AC	5068 - 5096
Universal Adapter	AATGATACGGCGACCACCGAGATCTACACGCCTCCCT CGCGCCATCAGAGATGTG	
Indexed Adapter**	CAAGCAGAAGACGGCATACGAGATNNNNNNGTGAC TGGAGTTCAGACGTGTGCTC	

*Underlined regions correspond to binding regions for Universal adaptor or Indexed Adaptor primers in 2nd round PCR reactions. Ns represent sites of random nucleotide unique Primer IDs for amplification from RNA (not applied in this study).

**In the indexed adaptor, the five Ns represent the unique index ID/barcode.

3.2 Amplifying HIV *env* V1V2 from 8E5 cells

3.2.1 PCR Round 1

Table A13. e*nv* V1V2 PCR Round 1 reaction components for 8E5 HIV genomes.

Component	**Volume per reaction (µl)**	**Stock concentration**	**Final concentration**
dH_20	21		
Forward primer V1F_UCT	1	10µM	0.2µM
Reverse primer (V2R_PID_UCT)	1	10µM	0.2µM
KAPA 2G Fast multiplex readymix	25	2x	1x
Template gDNA	2	0.075 - 0.000075 ng/µl	

Aliquot	48		
Total	50		

3.2.2 Indexing PCR

Table A14. Indexing PCR mix 1 components

Component	**Volume per reaction (µl)**	**Stock concentration**	**Final concentration**
dH_20	3.5-7.5		
dNTPs	1	10mM each	400µM
Universal adapter	1	10µM	0.4µM
Indexed adapter	1	10µM	0.4µM
Purified first round PCR product	2-6		
Mastermix aliquot	5.5 - 9.5		
Total	12.5		

Table A15. Indexing PCR mix 2 components

Component	**Volume per reaction (µl)**	**Stock concentration**	**Final concentration**
dH_20	9.63		
Expand high fidelity buffer	2.5	10x	1x
Expand high fidelity enzyme	0.38	3.5U/µl	2.6U/µl
Aliquot	12.5		
Total	12.5		

3.3 Restriction enzyme digestion of HIV from breastmilk and PBMC samples.

Table A16. Restriction enzyme digest reaction components.

Component	**Volume (µl)**
Template DNA	1-17.5 (maximum 1µg)
Cutsmart buffer (10x)	2
Ban II enzyme	0.5
mg H_20	0-16.5
Total	20

3.4 Amplifying breastmilk sample 8 and CAPRISA samples

3.4.1 PCR Round 1

Table A17. *env* V1V2 PCR Round 1 components for spiked-in and CAPRISA samples.

Component	**Volume per reaction (µl)**	**Stock concentration**	**Final concentration**
dH_20	3		
Forward primer V1F_UCT	1	10µM	0.2µM
Reverse primer V2R_PID_UCT	1	10µM	0.2µM
KAPA 2G Fast multiplex readymix	25	2x	1x
Digested template gDNA	20		
Aliquot	30		

Total	50		

Round 2 indexing PCR was carried as described in section 3.2.2.

3.5 Amplifying DolPHIN-2 samples

3.5.1 PCR Round 1 multiplexing (env and int)

Table A18. Multiplexed PCR Round 1 components.

Component	**Volume per reaction (µl)**	**Stock concentration**	**Final concentration**
mg H_20	1		
Forward primer V1F_UCT	1	10µM	0.2µM
Reverse primer V2R_PID_UCT	1	10µM	0.2µM
KAPA 2G Fast multiplex readymix	25	2x	1x
Int forward primer C9up_AD_UCT	1	10µM	0.2µM
Int cDNA primer C9dn_PID_UCT	1	10µM	0.2µM
Digested Template gDNA	20		
Aliquot	30		
Total	50		

Round 2 indexing PCR was carried as described in section 3.2.2.

3.6 Sequence analysis

Table A19. Script for dereplication of gDNA Illumina MiSeq sequences

Source	Script
Modified from https://github.com/HIVDiversity/clusteralign/blob/main/cluster_align.sh Developed by David Matten.	#/bin/bash # run this script like: bash <script_name>.sh < *.fasta from inside the folder where you want the output to go. # This lets people specify the files they wish to run into this code. mkdir -p 0_src 1_gapless 2_dereplicated 3_aligned for fn in $@ do echo ${fn} cp ${fn} 0_src/. cp ${fn} 1_gapless/. done cd 1_gapless sed -i 's/-//g' *.fasta for gapless_fn in `ls`
	do VSEARCH --derep_fulllength ${gapless_fn} --minuniquesize 2 --output ../2_dereplicated/${gapless_fn/.fasta}_derep.fasta done

Appendix 4: Sample integrity and HIV quantification

2.4 Processing and analysis of breastmilk samples 1 and 2, spiked with 8E5 cells only.

2.4.1 Breastmilk samples obtained from the breastmilk bank.

In order to determine the sensitivity of detection of low copy number HIV genomes in breastmilk using the ddPCR system, HIV naïve breastmilk samples from breastfeeding women were collected from Groote Schuur hospital for spiking with 8E5 cells. The cell count and viability of each breastmilk sample was measured using a Countess™ and cell viability recorded as shown in **Table A20**.

Table A20. Breastmilk cell count and viability of milk samples BM1, BM2 and BM3.

Breastmilk samples	**BM1**	**BM2**	**BM3**
Cell Viability (trypan blue) (%)	8	42	7
Concentration of breastmilk cells (live + dead) (cells/ml)	1.6×10^7	3.5×10^6	3.5×10^6

BM1 and BM3 demonstrated low viability (7% - 8%) and smaller cells, as expected for sterilized samples, however for sample 2 cells were bigger and had higher viability (42%). BM2 and BM3 both coincidentally had the same total cell count of 3.5×10^6 breastmilk cells/ml as compared to BM1 with 1.6×10^7 breastmilk cells/ml.

2.4.2 DNA extractions from breastmilk spiked with 8E5 cells only.

Table A21. NanoDrop results of breastmilk samples BM1 and BM2 following DNA extraction.

Samples (BM)	Sample concentration (ng/µl)	260/280 ratio	260/230 ratio
1A	10.54	1.32	0.64
1B	14.62	1.36	0.69
1C	4.06	1.63	0.72
1D	11.38	1.45	0.59
1E	1.41	1.56	0.43
1F	0.16	-3.68	-5.58
2A	8.84	1.22	0.79
2B	6.38	1.05	0.83
2C	5.69	1.05	0.87
2D	10.23	1.22	0.84
2E	6.34	1.14	1.05
2F	7.67	1.14	0.92

2.4.3 ddPCR of breastmilk samples spiked with 8E5 cells only

In order to assess the sensitivity of the ddPCR at detecting and quantifying HIV-1 DNA from 8E5 cells, HIV-naïve breastmilk samples BM1 and BM2 were spiked with 10-fold serially diluted 8E5 cells at concentrations ranging from 10 000 8E5 cells/million to 1 8E5 cell/million from samples A-F (based on cell counts in section 4.1.1 above). Genomic DNA from these samples were ddPCR amplified using *pol* and *rpp*30 primers and probes.

An amount of 20 ng of gDNA extracted from each breastmilk sample BM1 A-F and BM2 A-F was used for quantification of *pol* and *rpp*30 genes except for samples BM1E and BM1F whereby only 7.05ng and 0.8ng respectively were available. 8E5 cell gDNA at 20 ng input was used as a positive control and water non-template controls (NTCs) were included, as illustrated in **Figure A1**. ddPCR wells A-F contained breastmilk samples A-F respectively.

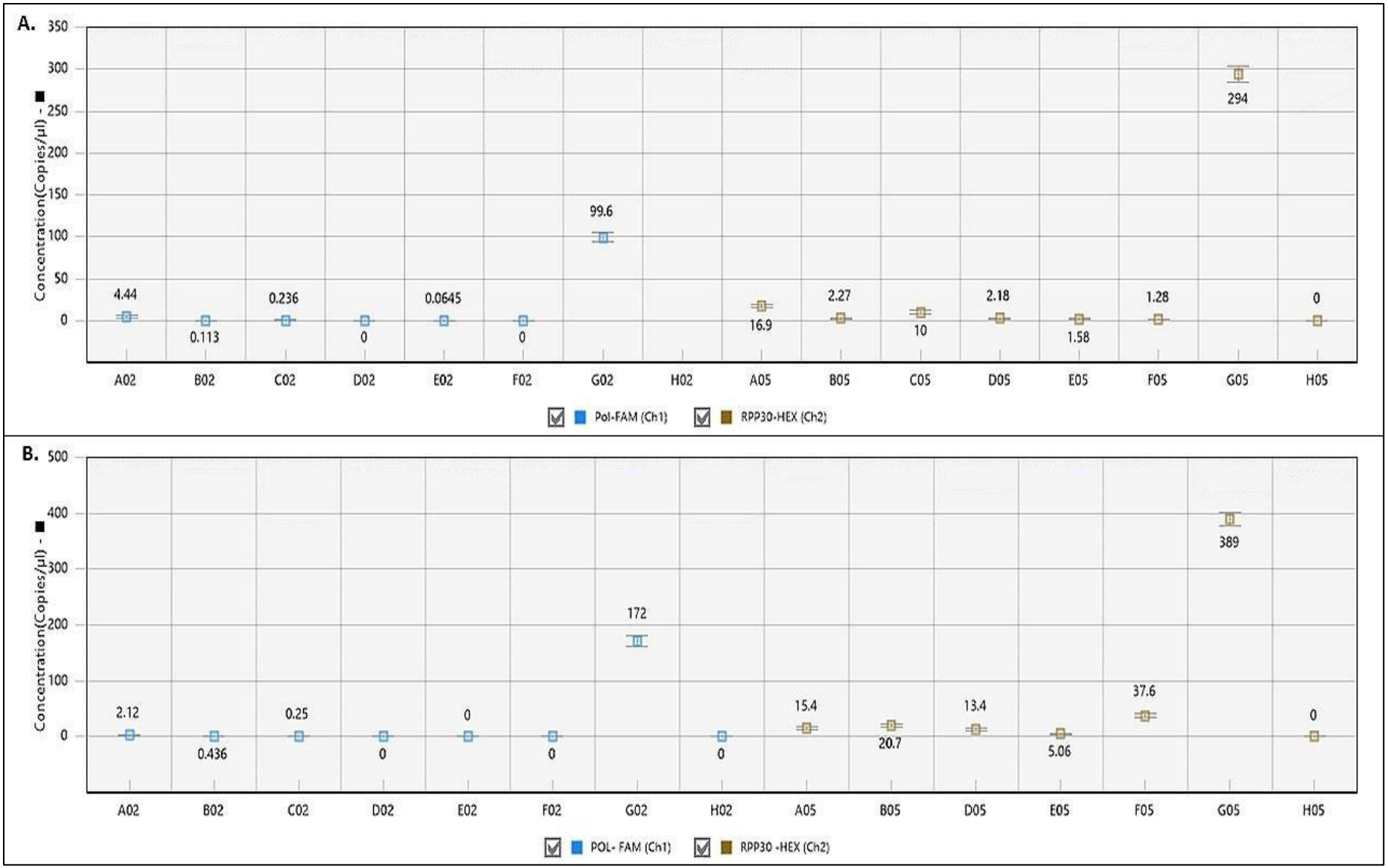
A.
Concentration(Copies/µl) - ■
350
300
250
200
150
100
50
0
4.44
0.113
0.236
0
0.0645
0
99.6
16.9
2.27
10
2.18
1.58
1.28
294
0
A02
B02
C02
D02
E02
F02
G02
H02
A05
B05
C05
D05
E05
F05
G05
H05
Pol-FAM (Ch1)
RPP30-HEX (Ch2)
B.
Concentration(Copies/µl) - ■
500
400
300
200
100
0
2.12
0.436
0.25
0
0
0
172
0
15.4
20.7
13.4
5.06
37.6
389
0
A02
B02
C02
D02
E02
F02
G02
H02
A05
B05
D05
E05
F05
G05
H05
POL- FAM (Ch1)
RPP30 -HEX (Ch2)

Figure A1. QuantaSoft generated graph showing concentration of *pol* (blue) and *rpp30* (brown) genes in breastmilk samples BM1 and BM2 following spiking with a range of 8E5 cells. A. Concentrations of BM1 A-F with gDNA input of 20 ng per reaction except for wells E and F with 7.05 ng and 0.8 ng inputs, respectively. B. Concentrations of BM2 with DNA input of 20 ng in wells A-F. All wells A02 to H02 show pol copy number concentrations while wells A05 to H05 indicate RPP30 copy number concentrations. Wells F contained non-spiked breastmilk samples. Wells G were positive controls containing 20 ng of 8E5 cell gDNA while wells H were water NTCs.

In comparison to the *pol* and *rpp30* positive controls, breastmilk samples BM1 and BM2 produced considerably lower DNA concentrations. Moreover, the *rpp30* concentrations were more than 10-fold lower than expected based on positive controls and varied across wells A05 to F05 whilst relatively uniform concentrations were expected for a standard cell number input. Breastmilk samples BM1B, BM1E and BM2C were excluded from further analysis since their concentrations were below the Poisson confidence maximum values for the experiment of 0.18 and 0.33 for BM1 and BM2, respectively. Hence, the concentrations obtained ranged from 0 to 4.44 *pol* copies/µl and 0 to 2.12 *pol* copies/µl and 8E5 HIV copy number ranged from 0 to 525 443.79 *pol* copies/million cells and 0 to 44 040.40 *pol* copies/million cells for BM1 and BM2 respectively. Since the highest spike-in value was 10 000 copies/million cells, these concentration were well above what was expected, likely due to low breastmilk cell number and poor DNA recovery (hence low rpp30 copies/ul) from these samples. *rpp30* copies were likely largely derived from the 8E5 cells since no background cells were spiked in. Regarding the 8E5 positive control, an average of 278 *rpp30* copies/ul was observed as well as an average of 135.8 *pol* copies/µl.

2.5 Processing and analysis of breastmilk samples BM4 and BM6, spiked with 8E5 cells and PBMCs.

2.5.1 DNA extractions from breastmilk spiked with 8E5 cells and PBMCs.

Table A22. NanoDrop results of breastmilk samples BM4 and BM6 following DNA extraction.

Samples (BM)	Sample concentration (ng/µl)	260/280 ratio	260/230 ratio
4A	34.24	1.69	1.45
4B	52.45	1.63	1.29
4C	51.65	1.71	1.43
4D	25.12	1.64	1.19
4E	29.05	1.62	1.19
4F	27.08	1.64	1.36
6A	127.13	1.45	1.56
6B	160.85	1.85	2.04
6C	66.41	1.79	1.94
6D	71.62	1.79	1.95
6E	108.77	1.86	2.13
6F	51.73	1.80	2.05

2.5.2 ddPCR of breastmilk samples spiked with 8E5 cells and PBMCs

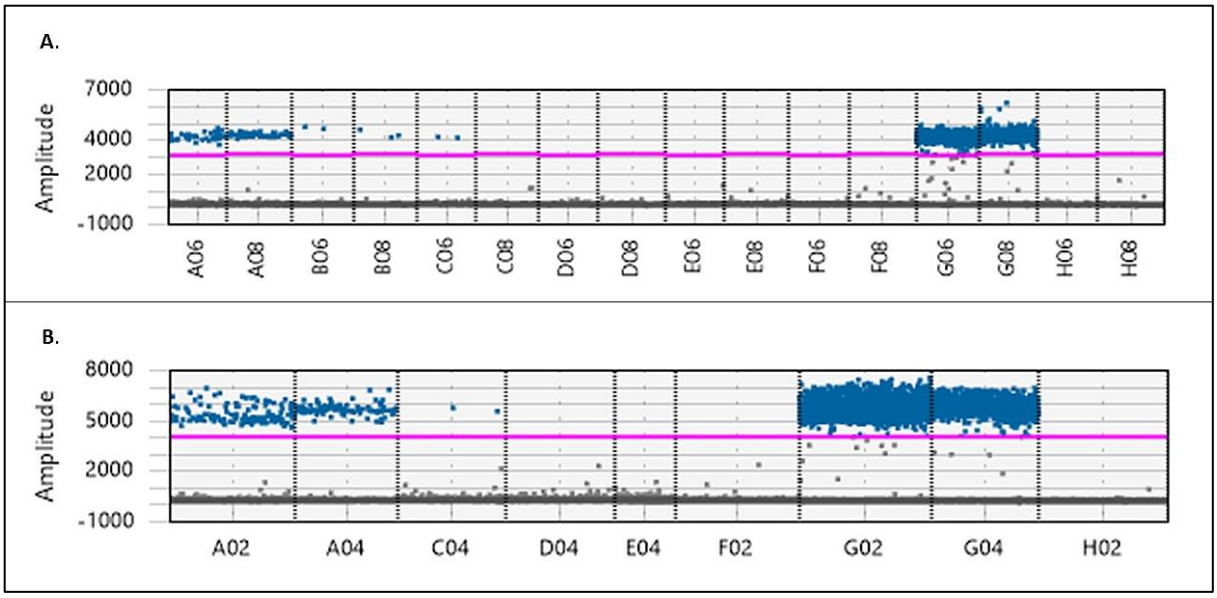

Figure A2. QuantaSoft™ 1-D amplitude plots of HIV *pol* (FAM) amplification signal of breastmilk samples BM4 and BM6 spiked with 10-fold serially diluted 8E5 cells at concentrations ranging from 10 000 8E5 cells/million to 1 8E5 cell/million from samples A-F. Positive droplets are in blue while negative droplets are in grey. **A**. BM4 A-F (DNA inputs ranging from 83.4ng to 262.3ng) illustrated in wells A06 to F06 respectively. This was duplicated in wells A08 to F08. **B**. BM6 A-F (DNA inputs ranging from 424.2ng to 1000.4ng) illustrated in wells A02 to F02 respectively. This was duplicated in wells A04 to F04. Wells containing less than 9000 droplets were excluded. Wells G contained 32.8ng of 8E5 positive control while wells H were water NTCs. The amplitude thresholds for exclusion of negative/background signals were set at 3211 for BM1 and 4052 for BM2 (pink lines).

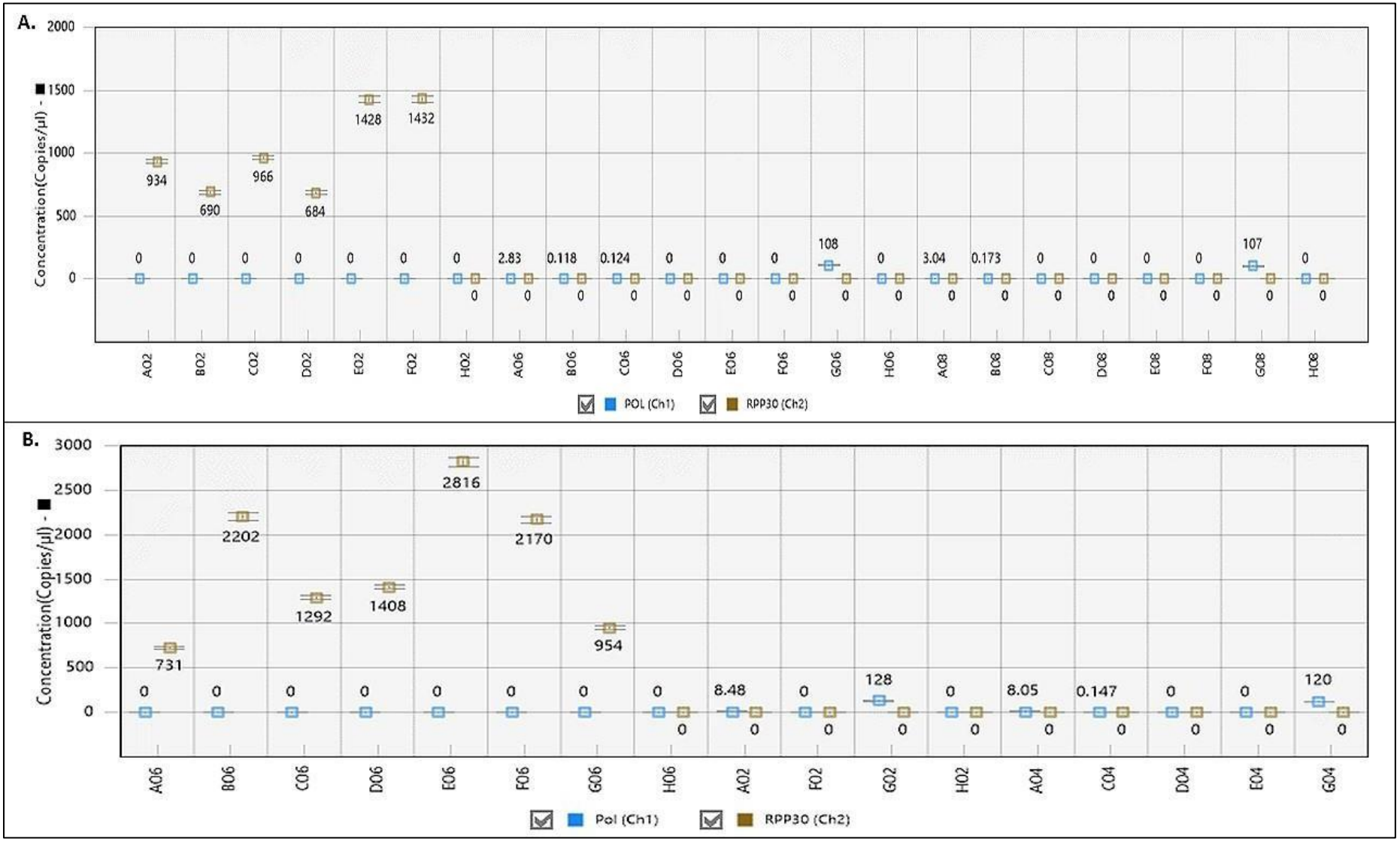
A.
Concentration(Copies/µl) - ■
2000
1500
1000
500
0
934
690
966
684
1428
1432
0
2.83
0.118
0.124
108
3.04
0.173
107
A02
B02
C02
D02
E02
F02
H02
A06
B06
C06
D06
E06
F06
G06
H06
A08
B08
C08
D08
E08
F08
G08
H08
POL (Ch1)
RPP30 (Ch2)
B.
Concentration(Copies/µl) - ■
3000
2500
2000
1500
1000
500
0
731
2202
1292
1408
2816
2170
954
8.48
128
8.05
0.147
120
A06
B06
C06
D06
E06
F06
G06
H06
A02
F02
G02
H02
A04
C04
D04
E04
G04
Pol (Ch1)
RPP30 (Ch2)

Figure A3. QuantaSoft generated graph showing concentration of *pol* (blue) and *rpp30* (brown) genes in breastmilk samples BM4 and BM6 following spiking with a range of 8E5 cells and PBMCs. A. shows BM4 whereby wells A02 to H02 show RPP30 concentrations. Wells A06 to H06 illustrate *pol* concentrations which were duplicated in wells A08 to H08. B. represent BM6 where wells A06 to H06 show *rpp30* concentrations and wells A02 to H02 indicate *pol* concentrations, duplicated in wells A04 to G04. All Wells F contained non-spiked breastmilk samples. All wells G were positive controls containing 8E5 cells and all wells H were water NTCs. Wells with less than 9000 droplets were excluded.

Table A23. Concentrations of HIV in breastmilk samples BM4 and BM6.

Samples (BM)	*pol* DNA input (ng)	Replicate 1 *pol* copies /µl	Replicate 1 with 1 µg DNA input *pol* copies /µl	Replicate 2 *pol* copies /µl	Replicate 2 with 1µg DNA input *pol* copies /µl	Average *pol* copies /µl with 1µg DNA input	*rpp30* DNA input (ng)	*rpp30* copies/µl	*rpp30* copies/ µL with 1ug DNA input	*pol* copies /million cells
4A	252.2	2.8	11.2	3.0	12.1	11.6	100.0	935.0	9350.0	2490.0
4B	86.3	0.0	0.0	0.0	0.0	0.0	52.7	690.0	13105.4	0.0
4C	273.9	0.0	0.0	0.0	0.0	0.0	100.0	966.0	9660.0	0.0
4D	83.4	0.0	0.0	0.0	0.0	0.0	50.9	685.0	13471.0	0.0
4E	262.3	0.0	0.0	0.0	0.0	0.0	100.0	1428.0	14280.0	0.0
4F	204.6	0.0	0.0	0.0	0.0	0.0	100.0	1434.0	14340.0	0.0
8E5	32.8	108.0	3292.7	107.0	3262.2	3277.4	100.0	0.0	0.0	0.0
NTC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
6A	1000.4	8.5	8.5	8.1	8.1	8.3	100.0	731.0	7310.0	2261.3
6B	1000.4	0.0	0.0	0.0	0.0	0.0	100.0	2202.0	22020.0	0.0
6C	544.6	0.0	0.0	0.1	0.3	0.0	100.0	1293.0	12930.0	0.0
6D	587.3	0.0	0.0	0.0	0.0	0.0	100.0	1408.0	14080.0	0.0
6E	891.9	0.0	0.0	0.0	0.0	0.0	100.0	2818.0	28180.0	0.0
6F	424.2	0.0	0.0	0.0	0.0	0.0	100.0	2170.0	21700.0	0.0
8E5	32.8	128.0	3902.4	120.0	3658.5	3780.5	100.0	954.0	9540.0	792555.1
NTC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

*Wells with concentrations below the Poisson confidence interval maximum were considered as background signal and assigned a value of zero.

2.6 Processing and analysis of breastmilk sample BM8, spiked with 8E5 cells and SUP-T1 cells

2.6.1 DNA extractions from breastmilk spiked with 8E5 cells and SUP-T1 cells

Table A24. NanoDrop results of breastmilk sample BM8 following DNA extraction.

Samples (BM)	NanoDrop sample concentration (ng/µl)	260/280 ratio	260/230 ratio
8A	757.28	1.95	2.09
8B	563.04	1.87	1.97
8C	739.58	1.98	2.16
8D	563.37	1.88	2.17
8E	665.78	1.97	2.05
8F	498.88	1.86	1.99

2.6.2 ddPCR of breastmilk samples spiked with 8E5 cells and SUP-T1 cells

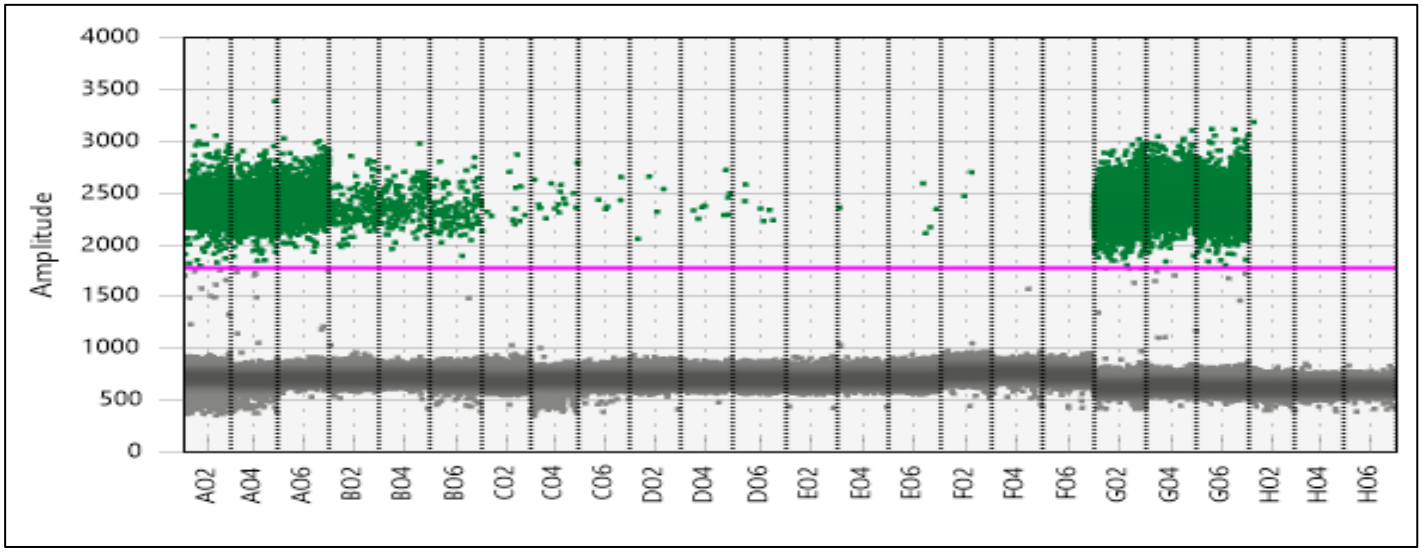

Figure A4. QuantaSoft™ 1-D amplitude plots of HIV *gag* (HEX) amplification signal of breastmilk sample BM8 spiked with 10-fold serially diluted 8E5 cells at concentrations ranging from 500 cells/µl to 0.05 cell/µl from samples A-F. Positive droplets are in green while negative droplets are in grey. BM8 A-F (1000ng DNA inputs) are shown in wells A02 to F02 respectively. This was triplicated in wells A04 to F04 and wells A06 to F06. Wells G contained 100ng of 8E5 positive control while wells H were water NTCs. The amplitude thresholds for exclusion of negative/background signals was set at 1778 (pink line).

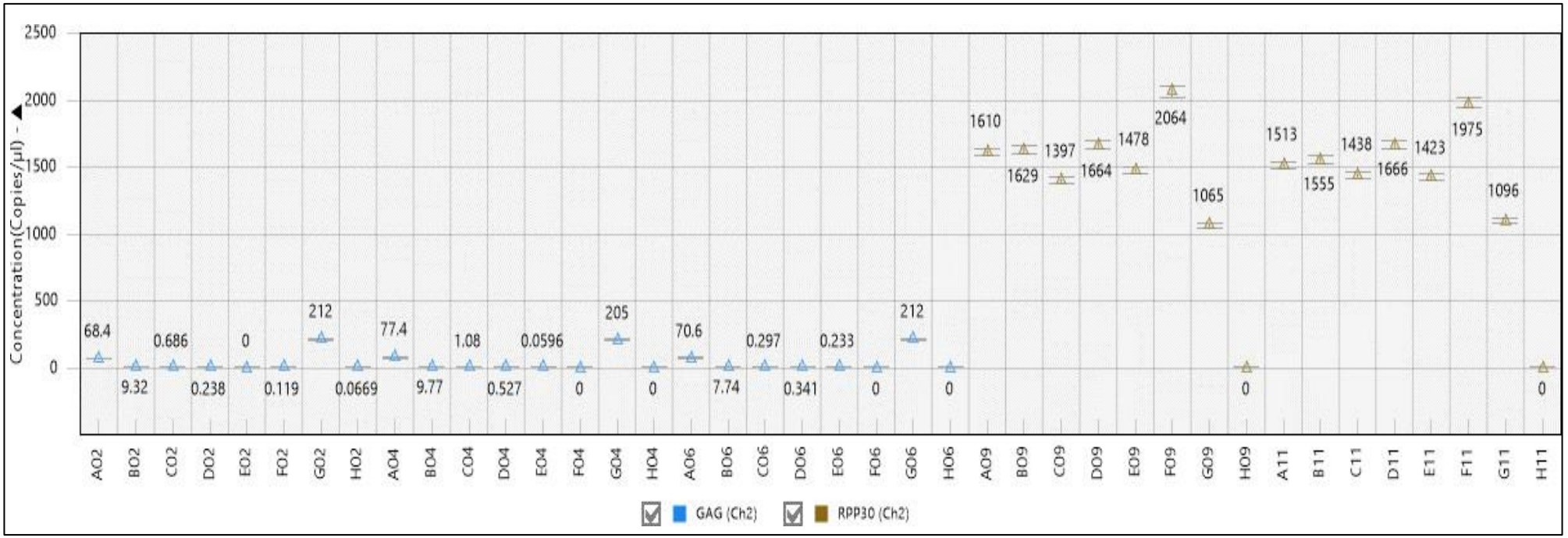

Figure A5. QuantaSoft generated graph showing concentration of gag (blue) and rpp30 (brown) genes in breastmilk sample BM8 following spiking with a range of 8E5 cells and SUP-T1 cells. Wells A02 to F02 show *gag* copy number concentrations from spiked samples A to F, replicated in wells A04 to F04 and A06 to F06. Wells A09 to H09 indicate *rpp30* concentrations for sample A to F as well as positive and NTC wells, duplicated in wells A11 to H11. All wells G were positive controls containing 8E5 cells and all wells H were water NTCs.

Table A25. Concentrations of breastmilk sample BM8.

Samples (BM)	*gag* DNA input (ng)	Replicate 1 *gag* copies/µl	Replicate 2 *gag* copies /µl	Replicate 3 *gag* copies /µl	Average *gag* copies /µl	*rpp30* DNA input (ng)	*rpp30* replicate 1 copies /µl	*rpp30* replicate 2 copies /µl	Average *rpp30* copies /µl	Average *rpp30* copies /µl with 1µg DNA input	*gag* copies / million cells
8A	1000.0	68.4	77.4	70.6	72.1	100.0	1609.9	1512.8	1561.3	15613.3	9240.0
8B	1000.0	9.3	9.8	7.7	8.9	100.0	1626.6	1555.2	1590.9	15909.0	1124.3
8C	1000.0	0.7	1.1	0.4	0.7	100.0	1396.7	1437.9	1417.3	14172.8	97.3
8D	1000.0	0.2	0.5	0.4	0.4	100.0	1663.3	1666.5	1664.9	16648.8	44.3
8E	1000.0	0.0	0.0	0.0	0.0	100.0	1478.2	1423.0	1450.6	14506.1	0.0
8F	1000.0	0.0	0.0	0.0	0.0	100.0	2050.6	1975.5	2013.0	20130.1	0.0
8E5	100.0	211.6	205.0	211.5	209.4	100.0	1064.6	1096.4	1080.5	N/A	387513.4
NTC	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0

2.7 Processing and analysis of DolPHIN-2 study samples

2.7.1 ddPCR of DolPHIN-2 study samples

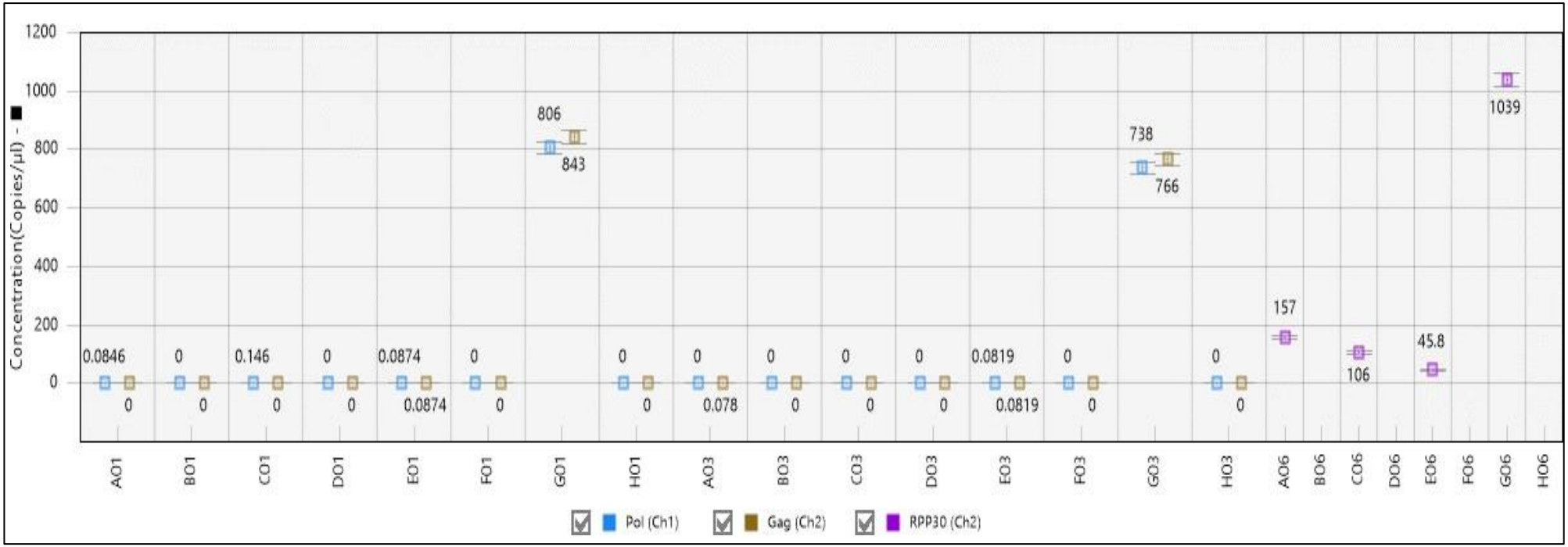

Figure A6. QuantaSoft generated graph showing concentrations of pol (blue), gag (brown) and rpp30 (purple) genes in DolPHIN-2 samples. Wells A01 to H01 show *pol* and *gag* concentrations, duplicated in wells A03 to H03 while wells A06 to H06 contained *rpp30* concentrations. Wells A, C, and E consist of samples 25008, 25054 and 25105 at 737.7 ng, 355.7 ng and 331.1 ng, respectively. Wells B, D, F and H consist of water NTCs while wells G include 8E5 positive controls at 847ng in wells G01 and G03 and 100 ng in well G06.

References

AAT Bioquest (2022) *What are the disadvantages of Nanodrop light spectrophotometer?* Available at: What are the disadvantages of Nanodrop light spectrophotometer? | AAT Bioquest (Accessed: 3 January 2024).

Abrahams, M.-R. *et al.* (2019) 'The replication-competent HIV-1 latent reservoir is primarily established near the time of therapy initiation', *Science Translational Medicine*, 11(513):eaaw5589. Available at: https://doi.org/10.1126/scitranslmed.aaw5589.

Acosta, R.K. *et al.* (2021) 'Three-year study of pre-existing drug resistance substitutions and efficacy of bictegravir/emtricitabine/tenofovir alafenamide in HIV-1 treatment-naive participants', *Journal of Antimicrobial Chemotherapy*, 76(8), pp. 2153–2157. Available at: https://doi.org/10.1093/jac/dkab115.

Altschul, S.F. *et al*. (1990) 'Basic local alignment search tool', *Journal of molecular biology,* 215(3), pp. 403-410. Available at: https://doi:10.1016/S0022-2836(05)80360-2.

American Type Culture Collection (ATCC) (2024) *8E5 [derivative of CEM].* Available at: https://www.atcc.org/products/crl-3617 (Accessed: 5 June 2023).

American Type Culture Collection (ATCC) (2024) *SUP-T1 [VB].* Available at: https://www.atcc.org/products/crl-1942 (Accessed: 5 June 2023).

Amin, O. *et al.* (2021) 'Understanding Viral and Immune Interplay During Vertical Transmission of HIV: Implications for Cure', *Frontiers in Immunology*, 12:757400. Available at: https://doi.org/10.3389/fimmu.2021.757400.

Ananworanich, J. *et al.* (2016) 'HIV DNA Set Point is Rapidly Established in Acute HIV Infection and Dramatically Reduced by Early ART', *EBioMedicine*, 11, pp. 68–72. Available at: https://doi.org/10.1016/j.ebiom.2016.07.024.

Anderson, E.M. and Maldarelli, F. (2018) 'Quantification of HIV DNA using droplet digital PCR techniques', *Current protocols in microbiology*, 51(1):62. Available at: https://doi.org/10.1002/cpmc.62.

Andreotti, M. *et al.* (2009) 'Comparison of HIV Type 1 Sequences from Plasma, Cell-Free Breast Milk, and Cell-Associated Breast Milk Viral Populations in Treated and Untreated Women in Mozambique', *AIDS Research and Human Retroviruses*, 25(7), pp. 707–711. Available at: https://doi.org/10.1089/aid.2008.0276.

Avettand-Fènoël, V. *et al.* (2016) 'Total HIV-1 DNA, a Marker of Viral Reservoir Dynamics with Clinical Implications', *Clinical Microbiology Reviews*, 29(4), pp. 859–880. Available at: https://doi.org/10.1128/CMR.00015-16.

Bäuerl, C. *et al.* (2022) 'SARS-CoV-2 RNA and antibody detection in breast milk from a prospective multicentre study in Spain', *Archives of Disease in Childhood - Fetal and Neonatal Edition*, 107(2), pp. 216–221. Available at: https://doi.org/10.1136/archdischild-2021-322463.

Becquart, P. *et al.* (2002) 'Compartmentalization of HIV-1 between Breast Milk and Blood of HIV-Infected Mothers', *Virology*, 300(1), pp. 109–117. Available at: https://doi.org/10.1006/viro.2002.1537.

Becquart, P. *et al.* (2006) 'Detection of a large T-cell reservoir able to replicate HIV-1 actively in breast milk', *AIDS*, 20(10), pp. 1453–1455. Available at: https://doi.org/10.1097/01.aids.0000233581.64467.55.

Becquart, P. *et al.* (2007) 'Diversity of HIV-1 RNA and DNA in breast milk from HIV-1-infected mothers', *Virology*, 363(2), pp. 256–260. Available at: https://doi.org/10.1016/j.virol.2007.02.003.

Belmonti, S., Di Giambenedetto, S. and Lombardi, F. (2021) 'Quantification of Total HIV DNA as a Marker to Measure Viral Reservoir: Methods and Potential Implications for Clinical Practice', *Diagnostics*, 12(1):39. Available at: https://doi.org/10.3390/diagnostics12010039.

Bergeron, L. and Sodroski, J. (1992) 'Dissociation of unintegrated viral DNA accumulation from single-cell lysis induced by human immunodeficiency virus type 1', *Journal of Virology*, 66(10), pp. 5777–5787. Available at: https://doi: 10.1128/JVI.66.10.5777-5787.1992.

BIO-RAD (2016) *QuantaSoft Analysis Pro*. Version 1.0. Bio-Rad, Hercules, California. Available at: https://www.bio-rad.com/en-za/product/qx200-droplet-digital-pcr-system?ID=MPOQQE4VY (Accessed: 1 December 2020).

BIO-RAD (no date) *Droplet Digital™ PCR Applications Guide*. Available at: https://www.bio-rad.com/webroot/web/pdf/lsr/literature/Bulletin_6407.pdf (Accessed: 11 March 2024).

Bispo, S. *et al.* (2017) 'Postnatal HIV transmission in breastfed infants of HIV-infected women on ART: a systematic review and meta-analysis', *Journal of the International AIDS Society*, 20(1):21251. Available at: https://doi.org/10.7448/IAS.20.1.21251.

Bobat, R et al. (1997) 'Breastfeeding by HIV-1-infected women and outcome in their infants: a cohort study from Durban, South Africa', *AIDS*, 11(13), pp. 1627-33. Available at: https://doi:10.1097/00002030-199713000-00012.

Bode, L. *et al.* (2014) 'It's Alive: Microbes and Cells in Human Milk and Their Potential Benefits to Mother and Infant', *Advances in Nutrition*, 5(5), pp. 571–573. Available at: https://doi.org/10.3945/an.114.006643.

Brodin, J. *et al.* (2016) 'Establishment and stability of the latent HIV-1 DNA reservoir', *eLife*, 5:e18889. Available at: https://doi.org/10.7554/eLife.18889.

Brownie, J. *et al.* (1997) 'The elimination of primer-dimer accumulation in PCR', *Nucleic Acids Research*, 25(16), pp. 3235–3241. Available at: https://doi: 10.1093/nar/25.16.3235.

Bruner, K.M. *et al.* (2016) 'Defective proviruses rapidly accumulate during acute HIV-1 infection', *Nature medicine*, 22(9), pp. 1043–1049. Available at: https://doi.org/10.1038/nm.4156.

Bunders, M.J. *et al.* (2012) 'Memory CD4+CCR5+ T cells are abundantly present in the gut of newborn infants to facilitate mother-to-child transmission of HIV-1', *Blood*, 120(22), pp. 4383–4390. Available at: https://doi.org/10.1182/blood-2012-06-437566.

Burnett, J.C. *et al.* (2010) 'Combinatorial Latency Reactivation for HIV-1 Subtypes and Variants', *Journal of Virology*, 84(12), pp. 5958–5974. Available at: https://doi.org/10.1128/JVI.00161-10.

Busby, E. *et al.* (2017) 'Instability of 8E5 calibration standard revealed by digital PCR risks inaccurate quantification of HIV DNA in clinical samples by qPCR', *Scientific Reports*, 7(1):1209. Available at: https://doi.org/10.1038/s41598-017-01221-5.

Caffrey, M. (2011) 'HIV envelope: challenges and opportunities for development of entry inhibitors' *Trends in microbiology,* 19(4), pp. 191-197. Available at: https://doi:10.1016/j.tim.2011.02.001.

Cervera, L. *et al.* (2019) 'Production of HIV-1-based virus-like particles for vaccination: achievements and limits', *Applied microbiology and biotechnology*, 103(18), pp. 7367-7384. Available at: https://doi:10.1007/s00253-019-10038-3.

Chan, C.N. *et al.* (2016) 'HIV-1 latency and virus production from unintegrated genomes following direct infection of resting CD4 T cells', *Retrovirology*, 13(1):1. Available at: https://doi.org/10.1186/s12977-015-0234-9.

Chen, B. (2019) 'Molecular Mechanism of HIV-1 Entry', *Trends in microbiology*, 27(10), pp. 878–891. Available at: https://doi.org/10.1016/j.tim.2019.06.002.

Chen, J. *et al.* (2022) 'The reservoir of latent HIV', *Frontiers in Cellular and Infection Microbiology*, 12:945956. Available at: https://doi.org/10.3389/fcimb.2022.945956.

Chun, T.-W. *et al.* (1995) 'In vivo fate of HIV-1-infected T cells: Quantitative analysis of the transition to stable latency', *Nature Medicine*, 1(12), pp. 1284–1290. Available at: https://doi.org/10.1038/nm1295-1284.

Chun, T.-W. *et al.* (1997) 'Quantification of latent tissue reservoirs and total body viral load in HIV-1 infection', *Nature*, 387(6629), pp. 183–188. Available at: https://doi.org/10.1038/387183a0.

Cohn, L.B., Chomont, N. and Deeks, S.G. (2020) 'The Biology of the HIV-1 Latent Reservoir and Implications for Cure Strategies', *Cell host & microbe*, 27(4), pp. 519–530. Available at: https://doi.org/10.1016/j.chom.2020.03.014.

Conduct Science (2024) *Automated Cell Counter.* Available at: Automated Cell Counter - Conduct Science (Accessed: 2 February 2024).

Crago, S.S. *et al.* (1979) 'Human colostral cells. I. Separation and characterization.', *Clinical and Experimental Immunology*, 38(3), pp. 585–597.

d'Ettorre, G *et al.* (2010) 'The role of HIV-DNA testing in clinical practice', *New Microbiologica,* 33, pp. 1-11.

Danaviah, S. *et al.* (2015) 'Evidence of Long-Lived Founder Virus in Mother-to-Child HIV Transmission', *PLOS ONE*, 10(3):e0120389. Available at: https://doi.org/10.1371/journal.pone.0120389.

Davis, N.L. *et al.* (2016) 'Maternal and Breast Milk Viral Load: Impacts of Adherence on Peri-Partum HIV Infections Averted - the BAN Study', *Journal of Acquired Immune Deficiency Syndromes*, 73(5), pp. 572–580. Available at: https://doi.org/10.1097/QAI.0000000000001145.

Davis, N.L. *et al.* (2019) 'Antiretroviral Drug Concentrations in Breastmilk, Maternal HIV Viral Load, and HIV Transmission to the Infant: Results From the BAN Study', *Journal of Acquired Immune Deficiency Syndromes*, 80(4), pp. 467–473. Available at: https://doi.org/10.1097/QAI.0000000000001941.

Deichmann, M. *et al.* (1997) 'Ultra-sensitive FISH is a useful tool for studying chronic HIV-1 infection', *Journal of virological methods,* 65(1), pp. 19-25. Available at: https://doi:10.1016/s0166-0934(96)02164-7.

Désiré, N. *et al.* (2001) 'Quantification of Human Immunodeficiency Virus Type 1 Proviral Load by a TaqMan Real-Time PCR Assay', *Journal of Clinical Microbiology*, 39(4), pp. 1303–1310. Available at: https://doi.org/10.1128/JCM.39.4.1303-1310.2001.

Dickinson, L. *et al.* (2021) 'Infant Exposure to Dolutegravir Through Placental and Breast Milk Transfer: A Population Pharmacokinetic Analysis of DolPHIN-1', *Clinical Infectious Diseases*, 73(5), pp. 1200–1207. Available at: https://doi.org/10.1093/cid/ciaa1861.

Dimitrov, D.S. *et al.* (1993) 'Quantitation of human immunodeficiency virus type 1 infection kinetics', *Journal of Virology*, 67(4), pp. 2182–2190. Available at: https://doi.org/10.1128/jvi.67.4.2182-2190.1993.

Doherty, T. *et al.* (2011) 'Implications of the new WHO guidelines on HIV and infant feeding for child survival in South Africa', *Bulletin of the World Health Organization*, 89(1), pp. 62–67. Available at: https://doi.org/10.2471/BLT.10.079798.

Dorosko, S.M. and Connor, R.I. (2010) 'Primary Human Mammary Epithelial Cells Endocytose HIV-1 and Facilitate Viral Infection of CD4+ T Lymphocytes', *Journal of Virology*, 84(20), pp. 10533–10542. Available at: https://doi.org/10.1128/JVI.01263-10.

Dube, S., Qin, J. and Ramakrishnan, R. (2008) 'Mathematical Analysis of Copy Number Variation in a DNA Sample Using Digital PCR on a Nanofluidic Device', *PLoS ONE*, 3(8):e2876. Available at: https://doi.org/10.1371/journal.pone.0002876.

Dyavar, S.R. *et al.* (2018) 'Normalization of cell associated antiretroviral drug concentrations with a novel RPP30 droplet digital PCR assay', *Scientific Reports*, 8 (1):3626. Available at: https://doi.org/10.1038/s41598-018-21882-0.

Esposito, A.M., Esposito, M.M. and Ptashnik, A. (2022) 'Phylogenetic Diversity of Animal Oral and Gastrointestinal Viromes Useful in Surveillance of Zoonoses', *Microorganisms*, 10(9):1815. Available at: https://doi.org/10.3390/microorganisms10091815.

Fanales-Belasio, E. *et al.* (2010) 'HIV virology and pathogenetic mechanisms of infection: a brief overview', *Annali dellIstituto Superiore di Sanità*, 46(1). Available at: https://doi.org/10.1590/S0021-25712010000100002.

Feldman-Winter, L. (2012) 'The AAP Updates Its Policy on Breastfeeding and Reaches Consensus on Recommended Duration of Exclusive Breastfeeding', *Journal of Human Lactation*, 28(2), pp. 116–117. Available at: https://doi.org/10.1177/0890334412442826.

Finzi, D. *et al.* (1997) 'Identification of a Reservoir for HIV-1 in Patients on Highly Active Antiretroviral Therapy', *Science*, 278(5341), pp. 1295–1300. Available at: https://doi.org/10.1126/science.278.5341.1295.

Finzi, D. *et al.* (1999) 'Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy', *Nature Medicine*, 5(5), pp. 512–517. Available at: https://doi.org/10.1038/8394.

Fish, C.S. *et al.* (2022) 'Protocol for high-throughput reservoir quantification across global HIV subtypes using a cross-subtype intact proviral DNA assay', *STAR Protocols*, 3(4):101681. Available at: https://doi.org/10.1016/j.xpro.2022.101681.

Fisher, R.G. *et al.* (2015) 'Next generation sequencing improves detection of drug resistance mutations in infants after PMTCT failure', *Journal of Clinical Virology*, 62, pp. 48–53. Available at: https://doi.org/10.1016/j.jcv.2014.11.014.

Flynn, P.M. *et al.* (2018) 'Prevention of HIV-1 Transmission Through Breastfeeding: Efficacy and Safety of Maternal Antiretroviral Therapy Versus Infant Nevirapine Prophylaxis for Duration of Breastfeeding in HIV-1-Infected Women With High CD4 Cell Count (IMPAACT PROMISE): A Randomized, Open-Label, Clinical Trial', *Journal of Acquired Immune Deficiency Syndromes*, 77(4), pp. 383–392. Available at: https://doi.org/10.1097/QAI.0000000000001612.

Flynn, P.M. *et al.* (2021) 'Association of Maternal Viral Load and CD4 Count With Perinatal HIV-1 Transmission Risk During Breastfeeding in the PROMISE Postpartum Component', *Journal of Acquired Immune Deficiency Syndromes*, 88(2), pp. 206–213. Available at: https://doi.org/10.1097/QAI.0000000000002744.

Folks, T. M. *et al.* (1986) 'Biological and biochemical characterization of a cloned Leu-3- cell surviving infection with the acquired immune deficiency syndrome retrovirus', *The Journal of experimental medicine,* 164(1), pp. 280-290. Available at: https://doi:10.1084/jem.164.1.280.

Fowler, M.G. *et al.* (2016) 'Benefits and Risks of Antiretroviral Therapy for Perinatal HIV Prevention', *New England Journal of Medicine*, 375(18), pp. 1726–1737. Available at: https://doi.org/10.1056/NEJMoa1511691.

Gantt, S. *et al.* (2010) 'Genetic Analyses of HIV-1 env Sequences Demonstrate Limited Compartmentalization in Breast Milk and Suggest Viral Replication within the Breast That Increases with Mastitis', *Journal of Virology*, 84(20), pp. 10812–10819. Available at: https://doi.org/10.1128/JVI.00543-10.

García, M. *et al.* (2018) 'Peering into the HIV reservoir', *Reviews in medical virology*, 28(4):e1981. Available at: https://doi.org/10.1002/rmv.1981.

Ghedira, R. *et al.* (2009) 'Assessment of Primer/Template Mismatch Effects on Real-Time PCR Amplification of Target Taxa for GMO Quantification', *Journal of Agricultural and Food Chemistry*, 57(20), pp. 9370–9377. Available at: https://doi.org/10.1021/jf901976a.

Ghosh, M.K. *et al.* (2003) 'Quantitation of Human Immunodeficiency Virus Type 1 in Breast Milk', *Journal of Clinical Microbiology*, 41(6), pp. 2465–2470. Available at: https://doi.org/10.1128/JCM.41.6.2465-2470.2003.

Goga, A.E. *et al.* (2021) 'Eliminating HIV transmission through breast milk from women taking antiretroviral drugs', *BMJ*, 374:n1697. Available at: https://doi.org/10.1136/bmj.n1697.

GraphPad Software (2007) *GraphPad Prism*. Version 5.00. GraphPad Software, San Diego, California. Available at: https://www.graphpad.com (Accessed: 14 August 2023).

Gray, R.R. *et al.* (2011) 'Multiple independent lineages of HIV-1 persist in breast milk and plasma', *AIDS*, 25(2), pp. 143–152. Available at: https://doi.org/10.1097/QAD.0b013e328340fdaf.

Groß, R. *et al.* (2020) 'Detection of SARS-CoV-2 in human breastmilk', *The Lancet*, 395(10239), pp. 1757–1758. Available at: https://doi.org/10.1016/S0140-6736(20)31181-8.

Haqshenas, G. *et al.* (2023) 'Development of a touchdown droplet digital PCR assay for the detection and quantitation of human papillomavirus 16 and 18 from self-collected anal samples', *Microbiology Spectrum*, 11(6):e01836-23. Available at: https://doi.org/10.1128/spectrum.01836-23.

Hassiotou, F. *et al.* (2013) 'Maternal and infant infections stimulate a rapid leukocyte response in breastmilk', *Clinical & translational immunology*, 2(4):e3. Available at:

https://doi:10.1038/cti.2013.1

Hassiotou, F. *et al.* (2013) 'Cells in human milk: state of the science', *Journal of human lactation : official journal of International Lactation Consultant Association*, 29(2), pp. 171-182. Available at: https://doi:10.1177/0890334413477242

Hatano, H. *et al.* (2013) 'Increase in 2–Long Terminal Repeat Circles and Decrease in D-dimer After Raltegravir Intensification in Patients With Treated HIV Infection: A Randomized, Placebo-Controlled Trial', *The Journal of Infectious Diseases*, 208(9), pp. 1436–1442. Available at: https://doi.org/10.1093/infdis/jit453.

Heath, L. *et al.* (2010) 'Restriction of HIV-1 Genotypes in Breast Milk Does Not Account for the Population Transmission Genetic Bottleneck That Occurs following Transmission', *PLoS ONE*, 5(4):e10213. Available at: https://doi.org/10.1371/journal.pone.0010213.

Henrich, T.J. *et al.* (2012) 'Low-level detection and quantitation of cellular HIV-1 DNA and 2-LTR circles using droplet digital PCR', *Journal of Virological Methods*, 186(1–2), pp. 68–72. Available at: https://doi.org/10.1016/j.jviromet.2012.08.019.

Her, C. and Weinshilboum, R.M. (1999) 'Endonuclease-Mediated Long PCR and Its Application to Restriction Mapping', *Current Issues in Molecular Biology*, 1(2), pp. 77-88. Available at: https://doi.org/10.21775/cimb.001.077.

Ho, Y.-C. *et al.* (2013) 'Replication-competent non-induced proviruses in the latent reservoir increase barrier to HIV-1 cure', *Cell*, 155(3), pp. 540–551. Available at: https://doi.org/10.1016/j.cell.2013.09.020.

Howison, M., Coetzer, M. and Kantor, R. (2018) 'Measurement error and variant-calling in deep Illumina sequencing of HIV', *Bioinformatics*, 35(12), pp. 2029–2035. Available at: https://doi.org/10.1093/bioinformatics/bty919.

Hu, T. *et al.* (2021) 'Next-generation sequencing technologies: An overview', *Human Immunology*, 82(11), pp. 801–811. Available at: https://doi.org/10.1016/j.humimm.2021.02.012.

Huang, K. *et al.* (2024) 'Exploring the Impact of Primer–Template Mismatches on PCR Performance of DNA Polymerases Varying in Proofreading Activity', *Genes*, 15(2):215.

Available at: https://doi.org/10.3390/genes15020215.

Illumina (2011) *Quality Scores for Next-Generation Sequencing.* Available at: https://www.illumina.com/documents/products/technotes/technote_Q-Scores.pdf (Accessed: 3 November 2023).

Illumina (2013) *Using a PhiX Control for HiSeq® Sequencing Runs.* Available at: https://www.illumina.com/content/dam/illumina-support/documents/products/technotes/technote_phixcontrolv3.pdf (Accessed: 22 May 2024).

Illumina (2017) *An Introduction to Next-Generation Sequencing Technology.* Available at: https://www.illumina.com/documents/products/illumina_sequencing_introduction.pdf (Accessed: 18 March 2024).

Integrated DNA Technologies (IDT) (2024) *OligoAnalyzer*. Available at: https://eu.idtdna.com/calc/analyzer (Accessed 10 May 2024).

Integrated DNA Technologies (IDT) (2024) *Frequently asked questions.* Available at: https://eu.idtdna.com/pages/support/faqs/how-do-i-use-the-oligoanalyzer-tool-to-analyze-possible-hairpins-and-dimers-formed-by-my-oligo (Accessed: 21 March 2024).

Invitrogen (2015) *Countess II Automated Cell Counters.* Available at: https://assets.thermofisher.com/TFS-Assets/BID/Reference-Materials/countess-II-automated-cell-counters-faq.pdf (Accessed: 5 January 2024).

Ishikawa, N. *et al.* (2016) 'Elimination of mother-to-child transmission of HIV and syphilis in Cuba and Thailand', *Bulletin of the World Health Organization*, 94(11), pp. 787-787A. Available at: https://doi.org/10.2471/BLT.16.185033.

Ismail, S.D. (2022) *Characterizing the cellular latent reservoir of HIV-1 and the effect of immune activation on characteristics of the reservoir.* PhD thesis. University of Cape Town, Cape Town.

Itiola, A.J. *et al.* (2019) 'Trends and predictors of mother-to-child transmission of HIV in an era of protocol changes: Findings from two large health facilities in North East Nigeria', *PLOS ONE*, 14(11):e0224670. Available at: https://doi.org/10.1371/journal.pone.0224670.

Jabara, C.B. *et al.* (2011) 'Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID', *Proceedings of the National Academy of Sciences of the United States of America*, 108(50):20166-71. Available at: https://doi:10.1073/pnas.1110064108.

John, G.C. *et al.* (2001) 'Correlates of Mother-to-Child Human Immunodeficiency Virus Type 1 (HIV-1) Transmission: Association with Maternal Plasma HIV-1 RNA Load, Genital HIV-1 DNA Shedding, and Breast Infections', *The Journal of Infectious Diseases*, 183(2), pp. 206–212. Available at: https://doi.org/10.1086/317918.

John-Stewart, G. *et al.* (2004) 'Breast-feeding and Transmission of HIV-1', *Journal of Acquired Immune Deficiency Syndromes*, 35(2), pp. 196–202. Available at: https://doi.org/10.1097/00126334-200402010-00015.

Joseph, S.B. *et al.* (2024) 'The timing of HIV-1 infection of cells that persist on therapy is not strongly influenced by replication competency or cellular tropism of the provirus', *PLoS Pathogens*, 20(2):e1011974. Available at: https://https://doi.org/ 10.1371/journal.ppat.1011974.

Kantarci, S. *et al.* (2007) 'Subclinical Mastitis, Cell-Associated HIV-1 Shedding in Breast Milk, and Breast-Feeding Transmission of HIV-1', *Journal of Acquired Immune Deficiency Syndromes*, 46(5), pp. 651–654. Available at: https://doi.org/10.1097/QAI.0b013e31815b2db2.

Kesho Bora Study Group (2011) 'Triple antiretroviral compared with Zidovudine and single-dose Nevirapine prophylaxis during pregnancy and breastfeeding for prevention of mother-to-child transmission of HIV-1 (Kesho Bora study): a randomised controlled trial', *The Lancet Infectious diseases*, 11(3), pp. 171-80. Available at: https://doi.org/10.1016/S1473-3099(10)70288-7.

Kintu, K. *et al.* (2020) 'Dolutegravir versus Efavirenz in women starting HIV therapy in late pregnancy (DolPHIN-2): an open-label, randomised controlled trial', *The Lancet HIV*, 7(5), pp. 332–339. Available at: https://doi.org/10.1016/S2352-3018(20)30050-3.

Kojabad, A.A. *et al.* (2021) 'Droplet digital PCR of viral DNA/RNA, current progress, challenges, and future perspectives', *Journal of Medical Virology*, 93(7), pp. 4182–4197. Available at:

https://doi.org/10.1002/jmv.26846.

Koulinska, Irene N *et al.* (2006) 'Risk of HIV-1 transmission by breastfeeding among mothers infected with recombinant and non-recombinant HIV-1 genotypes', *Virus Research*, 120(1–2), pp. 191–198. Available at: https://doi.org/10.1016/j.virusres.2006.03.007.

Koulinska, Irene N *et al.* (2006) 'Transmission of Cell-Free and Cell-Associated HIV-1 Through Breast-Feeding', *Journal of Acquired Immune Deficiency Syndromes*, 41(1), pp. 93–99. Available at: https://doi.org/10.1097/01.qai.0000179424.19413.24.

Kourtis, A.P. and Bulterys, M. (2010) 'Mother-to-Child Transmission of HIV: Pathogenesis, Mechanisms and Pathways', *Clinics in Perinatology*, 37(4), pp. 721–737. Available at: https://doi.org/10.1016/j.clp.2010.08.004.

Kuhn, L. *et al.* (2013) 'HIV-1 concentrations in human breast milk before and after weaning', *Science translational medicine*, 5(181):181ra51. Available at: https://doi.org/10.1126/scitranslmed.3005113.

Kumar, S. *et al,* (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across computing platforms' *Molecular Biology and Evolution,* 35(6), pp. 1547-1549. Available at: https://doi: 10.1093/molbev/msy096.

Kumar, S. *et al.* (2018) 'MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms', *Molecular Biology and Evolution*, 35(6), pp. 1547–1549. Available at: https://doi.org/10.1093/molbev/msy096.

Labuschagne, J.P.L. (2018) *Development of a data processing toolkit for the analysis of next-generation sequencing data generated using the primer ID approach.* PhD thesis. University of Western Cape, Cape Town.

Lallemant, M. *et al.* (2020) 'Perinatal Antiretroviral Intensification to Prevent Intrapartum HIV Transmission When Antenatal Antiretroviral Therapy Is Initiated Less Than 8 Weeks Before Delivery', *Journal of acquired immune deficiency syndromes*, 84(3), pp. 313–322. Available at: https://doi.org/10.1097/QAI.0000000000002350.

Larsson, A. (2014) 'AliView: a fast and lightweight alignment viewer and editor for large datasets', *Bioinformatics*, 30(22), pp. 3276–3278. Available at: https://doi.org/10.1093/bioinformatics/btu531.

Lawrence, R.A., and Lawrence, R.M. (1999) 'Physiology of lactation' in R. A. Lawrence and R. M. Lawrence (ed.), *Breastfeeding: a guide for the medical profession*. Mosby, Boston, MA, pp. 59–94.

Lehman, D.A. *et al.* (2008) 'HIV-1 persists in breast milk cells despite antiretroviral treatment to prevent mother-to-child transmission', *AIDS*, 22(12), pp. 1475–1485. Available at: https://doi.org/10.1097/QAD.0b013e328302cc11.

Li, L. (2001) 'Role of the non-homologous DNA end joining pathway in the early steps of retroviral infection', *The EMBO Journal*, 20(12), pp. 3272–3281. Available at: https://doi.org/10.1093/emboj/20.12.3272.

Lindegren, M.L. (1999) 'Trends in Perinatal Transmission of HIV/AIDS in the United States', *JAMA*, 282(6):531. Available at: https://doi.org/10.1001/jama.282.6.531.

Lockman, S. *et al.* (2021) 'Efficacy and safety of Dolutegravir- and tenofovir alafenamide fumarate-containing HIV antiretroviral treatment regimens started in pregnancy: a randomised controlled trial', *Lancet*, 397(10281), pp. 1276–1292. Available at: https://doi.org/10.1016/S0140-6736(21)00314-7.

Los Alamos National Laboratory (LANL) (2014) *Hypermut.* Available at: https://www.hiv.lanl.gov/content/sequence/HYPERMUT/hypermut.html (Accessed 17 January 2023).

Los Alamos National Laboratory (LANL) (2019) *HIV Sequence Locator.* Available at: https://www.hiv.lanl.gov/content/sequence/LOCATE/locate.html (Accessed: 15 April 2023).

Los Alamos National Laboratory (LANL) (2020) *AnalyzeAlign.* Available at: https://www.hiv.lanl.gov/content/sequence/ANALYZEALIGN/analyze_align.html (Accessed: 4 May 2023).

Lunney, K.M. *et al.* (2010) 'Associations between Breast Milk Viral Load, Mastitis, Exclusive Breast-Feeding, and Postnatal Transmission of HIV', *Clinical Infectious Diseases*, 50, pp. 762–769. Available at: https://doi.org/10.1086/650535. 50:762–769.

Luzuriaga, K. and Mofenson, L.M. (2016) 'Challenges in the Elimination of Pediatric HIV-1 Infection', *New England Journal of Medicine*, 374(8), pp. 761–770. Available at: https://doi.org/10.1056/NEJMra1505256.

Lyons, K.E. *et al.* (2020) 'Breast Milk, a Source of Beneficial Microbes and Associated Benefits for Infant Health', *Nutrients*, 12(4):1039. Available at: https://doi.org/10.3390/nu12041039.

Malaba, T.R. *et al.* (2022) '72 weeks post-partum follow-up of Dolutegravir versus Efavirenz initiated in late pregnancy (DolPHIN-2): an open-label, randomised controlled study', *The Lancet HIV*, 9(8), pp. 534–543. Available at: https://doi.org/10.1016/S2352-3018(22)00173-4.

Mangold, J.F. *et al.* (2021) 'Maternal Interventions to Prevent Mother-To-Child Transmission of HIV: Moving Beyond Antiretroviral Therapy', *The Pediatric infectious disease journal*, 40(5 Suppl), pp. S5–S10. Available at: https://doi.org/10.1097/INF.0000000000002774.

Martinez-Picado, J. *et al.* (2018) 'Episomal HIV-1 DNA and its relationship to other markers of HIV-1 persistence', *Retrovirology*, 15(1):15. Available at: https://doi.org/10.1186/s12977-018-0398-1.

Masago, K. *et al.* (2021) 'Comparison between Fluorimetry (Qubit) and Spectrophotometry (NanoDrop) in the Quantification of DNA and RNA Extracted from Frozen and FFPE Tissues from Lung Cancer Patients: A Real-World Use of Genomic Tests', *Medicina*, 57(12):1375. Available at: https://doi.org/10.3390/medicina57121375.

Masuda, T. (2011) 'Non-Enzymatic Functions of Retroviral Integrase: The Next Target for Novel Anti-HIV Drug Development', *Frontiers in microbiology*, 13(2):210. Available at: https://doi:10.3389/fmicb.2011.00210.

Mbonye, U. and Karn, J. (2017) 'The Molecular Basis for Human Immunodeficiency Virus Latency', *Annual Review of Virology*, 4(1), pp. 261–285. Available at: https://doi.org/10.1146/annurev-virology-101416-041646.

McGuire, M.K. *et al.* (2021) 'Best Practices for Human Milk Collection for COVID-19 Research', *Breastfeeding Medicine*, 16(1), pp. 29–38. Available at: https://doi.org/10.1089/bfm.2020.0296.

Meissner, M.E., Talledge, N. and Mansky, L.M. (2022) 'Molecular Biology and Diversification of Human Retroviruses', *Frontiers in virology*, 2:872599. Available at: https://doi.org/10.3389/fviro.2022.872599.

Milligan, C. and Overbaugh, J. (2014) 'The Role of Cell-Associated Virus in Mother-to-Child HIV Transmission', *The Journal of Infectious Diseases*, 210(3), pp. 631–640. Available at: https://doi.org/10.1093/infdis/jiu344.

miniPCR bio (no date) *Troubleshooting primer dimer in PCR.* Available at: https://www.minipcr.com/primer-dimer-pcr/ (Accessed: 10 December 2023).

Mukherjee, S. *et al.* (2015) 'Large-scale contamination of microbial isolate genomes by Illumina PhiX control', *Standards in Genomic Sciences*, 10(1):18. Available at: https://doi.org/10.1186/1944-3277-10-18.

Munyuza, C., Ji, H. and Lee, E.R. (2022) 'Probe Capture Enrichment Methods for HIV and HCV Genome Sequencing and Drug Resistance Genotyping', *Pathogens*, 11(6):693. Available at: https://doi.org/10.3390/pathogens11060693

Myer, L. *et al.* (2017) 'HIV viraemia and mother-to-child transmission risk after antiretroviral therapy initiation in pregnancy in Cape Town, South Africa', *HIV Medicine*, 18(2), pp. 80–88. Available at: https://doi.org/10.1111/hiv.12397.

Nakamura, K.J. *et al.* (2017) 'Breast milk and in utero transmission of HIV-1 select for envelope variants with unique molecular signatures', *Retrovirology*, 14(1):6. Available at: https://doi.org/10.1186/s12977-017-0331-z.

National Center for Biotechnology Information (NCBI) (no date) *BLAST: Basic Local Alignment Search Tool*. Available at: https://blast.ncbi.nlm.nih.gov/Blast.cgi?PROGRAM=blastn&PAGE_TYPE=BlastSearch&LINK_LOC=blasthome (Accessed: 21 June 2024).

Ndirangu, J. *et al.* (2012) 'Cell-Free (RNA) and Cell-Associated (DNA) HIV-1 and Postnatal Transmission through Breastfeeding', *PLoS ONE*, 7(12):e51493. Available at: https://doi.org/10.1371/journal.pone.0051493.

Ngandu, N.K. *et al.* (2022) 'HIV viral load non-suppression and associated factors among pregnant and postpartum women in rural northeastern South Africa: a cross-sectional survey', *BMJ Open*, 12(3):e058347. Available at: https://doi.org/10.1136/bmjopen-2021-058347.

Njom Nlend, A.E *et al.* (2018) 'HIV-1 transmission and survival according to feeding options in infants born to HIV-infected women in Yaoundé, Cameroon', *BMC pediatrics,* 18(1):69. Available at: https://doi:10.1186/s12887-018-1049-3.

Njom Nlend, A.E. (2022) 'Mother-to-Child Transmission of HIV Through Breastfeeding Improving Awareness and Education: A Short Narrative Review', *International Journal of Women's Health*, 14, pp. 697–703. Available at: https://doi.org/10.2147/IJWH.S330715.

Notarbartolo, V. *et al.* (2022) 'Composition of Human Breast Milk Microbiota and Its Role in Children's Health', *Pediatric Gastroenterology, Hepatology & Nutrition*, 25(3), pp. 194–210. Available at: https://doi.org/10.5223/pghn.2022.25.3.194.

Ouyang, F. *et al.* (2024) 'HIV-1 Drug Resistance Detected by Next-Generation Sequencing among ART-Naïve Individuals: A Systematic Review and Meta-Analysis', *Viruses*, 16(2):239. Available at: https://doi.org/10.3390/v16020239.

Pace, M.J., Graf, E.H. and O'Doherty, U. (2013) 'HIV 2-long terminal repeat circular DNA is stable in primary CD4+T Cells', *Virology*, 441(1), pp. 18–21. Available at: https://doi.org/10.1016/j.virol.2013.02.028.

Pan, R. *et al.* (2015) 'The V1V2 Region of HIV-1 gp120 Forms a Five-Stranded Beta Barrel', *Journal of Virology*, 89(15), pp. 8003–8010. Available at: https://doi.org/10.1128/JVI.00754-15.

Pankau, M.D. *et al.* (2020) 'Dynamics of HIV DNA reservoir seeding in a cohort of superinfected Kenyan women', *PLOS Pathogens*, 16(2):e1008286. Available at: https://doi.org/10.1371/journal.ppat.1008286.

Perelson, A.S. *et al.* (1996) 'HIV-1 Dynamics in Vivo: Virion Clearance Rate, Infected Cell Life-Span, and Viral Generation Time', *Science*, 271(5255), pp. 1582–1586. Available at: https://doi.org/10.1126/science.271.5255.1582.

Persson, S. *et al.* (2019) 'Missing the Match Might Not Cost You the Game: Primer-Template Mismatches Studied in Different Hepatitis A Virus Variants', *Food and Environmental Virology*, 11(3), pp. 297–308. Available at: https://doi.org/10.1007/s12560-019-09387-z.

Petitjean, G. *et al.* (2007) 'Isolation and characterization of HIV-1-infected resting CD4+ T lymphocytes in breast milk', *Journal of Clinical Virology*, 39(1), pp. 1–8. Available at: https://doi.org/10.1016/j.jcv.2007.02.004.

Powell DM, et al. (1988) 'Cell line producing AIDS viral antigens without producing infectious virus particles', US Patent 4,752,565.

Prediger, E. (2013) *How to design primers and probes for PCR and qPCR.* Available at: https://eu.idtdna.com/pages/education/decoded/article/designing-pcr-primers-and-probes (Accessed: 21 February 2023).

Prendergast, AJ. *et al.* (2019) 'Transmission of CMV, HTLV-1, and HIV through breastmilk', *Lancet Child Adolescent Health*, 3(4), pp. 264-273. Available at https://doi: 10.1016/S2352-4642(19)30024-0.

QIAGEN (2020) *MinElute® Handbook.* Available at: MinElute Handbook - QIAGEN (Accessed: 5 October 2023).

QIAGEN (2020) *AllPrep DNA/RNA Mini Handbook.* Available at: AllPrep DNA/RNA Mini Handbook - QIAGEN (Accessed: 5 October 2023)

Ramdas, P. *et al.* (2020) 'From Entry to Egress: Strategic Exploitation of the Cellular Processes by HIV-1', *Frontiers in Microbiology*, 11:559792. Available at: https://doi.org/10.3389/fmicb.2020.559792.

Re, M.C. *et al.* (2010) 'HIV-1 DNA proviral load in treated and untreated HIV-1 sero-positive patients', *Clinical Microbiology and Infection*, 16(6), pp. 640–646. Available at: https://doi.org/10.1111/j.1469-0691.2009.02826.x.

Read, J.S. (2003) 'Human Milk, Breastfeeding, and Transmission of Human Immunodeficiency Virus Type 1 in the United States', *Pediatrics,* 112(5), pp. 1196–1205. Available at: https://doi.org/10.1542/peds.112.5.1196.

Richardson, B.A. *et al.* (2003) 'Breast-Milk Infectivity in Human Immunodeficiency Virus Type 1–Infected Mothers', *The Journal of Infectious Diseases*, 187(5), pp. 736–740. Available at: https://doi.org/10.1086/374272.

Rognes, T. *et al.* (2016) 'VSEARCH: a versatile open source tool for metagenomics', *PeerJ*, 4:e2584. Available at: https://doi.org/10.7717/peerj.2584.

Rousseau, C.M. *et al.* (2003) 'Longitudinal Analysis of Human Immunodeficiency Virus Type 1 RNA in Breast Milk and of Its Relationship to Infant Infection and Maternal Disease', *The Journal of Infectious Diseases*, 187(5), pp. 741–747. Available at: https://doi.org/10.1086/374273.

Rowlands, V. *et al.* (2019) 'Optimisation of robust singleplex and multiplex droplet digital PCR assays for high confidence mutation detection in circulating tumour DNA', *Scientific Reports*, 9(1):12620. Available at: https://doi.org/10.1038/s41598-019-49043-x.

Rutagwera, D.G. *et al.* (2022) 'Recurrent Severe Subclinical Mastitis and the Risk of HIV Transmission Through Breastfeeding', *Frontiers in Immunology*, 13:822076. Available at: https://doi.org/10.3389/fimmu.2022.822076.

Rutsaert, S. *et al.* (2018) 'Digital PCR as a tool to measure HIV persistence', *Retrovirology*, 15:16. Available at: https://doi.org/10.1186/s12977-018-0399-0.

Sah, S. *et al.* (2013) 'Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies', *Genome Medicine*, 5(8):77. Available at: https://doi.org/10.1186/gm481.

Salazar-Gonzalez, J.F. *et al.* (2011) 'Origin and Evolution of HIV-1 in Breast Milk Determined by Single-Genome Amplification and Sequencing', *Journal of Virology*, 85(6), pp. 2751–2763. Available at: https://doi.org/10.1128/JVI.02316-10.

Samuel, R. *et al.* (2014) 'Minority HIV-1 Drug-Resistant Mutations and Prevention of Mother-to-Child Transmission: Perspectives for Resource-Limited Countries', *AIDS Reviews,* 16(4), pp. 187-198.

Satomi, M. *et al.* (2005) 'Transmission of Macrophage-Tropic HIV-1 by Breast-Milk Macrophages via DC-SIGN', *The Journal of Infectious Diseases*, 191(2), pp. 174–181. Available at: https://doi.org/10.1086/426829.

Semba, R.D. *et al.* (1999) 'Human Immunodeficiency Virus Load in Breast Milk, Mastitis, and Mother-to-Child Transmission of Human Immunodeficiency Virus Type 1', *The Journal of Infectious Diseases*, 180(1), pp. 93–98. Available at: https://doi.org/10.1086/314854.

Semrau, K. *et al.* (2008) 'Temporal and Lateral Dynamics of HIV Shedding and Elevated Sodium in Breast Milk Among HIV-Positive Mothers During the First 4 Months of Breast-Feeding', *Journal of Acquired Immune Deficiency Syndromes*, 47(3), pp. 320–328. Available at: https://doi.org/10.1097/QAI.0b013e31815e7436.

Shapiro, R.L. *et al.* (2005) 'Highly Active Antiretroviral Therapy Started during Pregnancy or Postpartum Suppresses HIV-1 RNA, but Not DNA, in Breast Milk', *The Journal of Infectious Diseases*, 192(5), pp. 713–719. Available at: https://doi.org/10.1086/432489.

Siliciano, J.D. and Siliciano, R.F. (2022) 'In Vivo Dynamics of the Latent Reservoir for HIV-1: New Insights and Implications for Cure', *Annual Review of Pathology: Mechanisms of Disease*, 17(1), pp. 271–294. Available at: https://doi.org/10.1146/annurev-pathol-050520-112001.

Siliciano, J.D. *et al.* (2003) 'Long-term follow-up studies confirm the stability of the latent reservoir for HIV-1 in resting CD4+ T cells', *Nature Medicine*, 9(6), pp. 727–728. Available at: https://doi.org/10.1038/nm880.

Simbolo, M. *et al.* (2013) 'DNA Qualification Workflow for Next Generation Sequencing of Histopathological Samples', *PLoS ONE*, 8(6):e62692. Available at: https://doi.org/10.1371/journal.pone.0062692.

Slyker, J.A. *et al.* (2012) 'Incidence and Correlates of HIV-1 RNA Detection in the Breast Milk of Women Receiving HAART for the Prevention of HIV-1 Transmission', *PLoS ONE*, 7(1):e29777. Available at: https://doi.org/10.1371/journal.pone.0029777.

Smith, S D *et al.* (1988) 'Clinical and biologic characterization of T-cell neoplasias with rearrangements of chromosome 7 band q34', *Blood,* 71(2), pp. 395-402.

Spina, C.A. *et al.* (2013) 'An In-Depth Comparison of Latent HIV-1 Reactivation in Multiple Cell Model Systems and Resting CD4+ T Cells from Aviremic Patients', *PLoS Pathogens*, 9(12):e1003834. Available at: https://doi.org/10.1371/journal.ppat.1003834.

Strain, M.C. *et al.* (2013) 'Highly Precise Measurement of HIV DNA by Droplet Digital PCR', *PLoS ONE*, 8(4):e55943. Available at: https://doi.org/10.1371/journal.pone.0055943.

Süß, B. *et al.* (2009) 'Studying the effect of single mismatches in primer and probe binding regions on amplification curves and quantification in real-time PCR', *Journal of Microbiological Methods*, 76(3), pp. 316–319. Available at: https://doi.org/10.1016/j.mimet.2008.12.003.

Svicher, V. *et al.* (2014) 'Understanding HIV Compartments and Reservoirs', *Current HIV/AIDS Reports*, 11(2), pp. 186–194. Available at: https://doi.org/10.1007/s11904-014-0207-y.

Tavassoli, A. (2011) 'Targeting the protein–protein interactions of the HIV lifecycle', *Chemical Society Reviews*, 40(3), pp. 1337–1346. Available at: https://doi.org/10.1039/C0CS00092B.

The Human Protein Atlas (2022) *RPP30.* Available at: RPP30 protein expression summary - The Human Protein Atlas (Accessed: 12 January 2024).

Thermo Fisher Scientific (2016) *NanoDrop One User Guide.* Available at: https://tools.thermofisher.com/content/sfs/manuals/3091-NanoDrop-One-Help-UG-en.pdf (Accessed: 2 March 2024).

Thermo Fisher Scientific (no date) *T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers 260/280 and 260/230 Ratios.* Available at: https://dna.uga.edu/wp-content/uploads/sites/51/2019/02/Note-on-the-260_280-and-260_230-Ratios.pdf (Accessed: 30 March 2024).

Thomas, T.K. *et al.* (2011) 'Triple-Antiretroviral Prophylaxis to Prevent Mother-To-Child HIV Transmission through Breastfeeding—The Kisumu Breastfeeding Study, Kenya: A Clinical Trial', *PLoS Medicine*, 8(3):e1001015. Available at: https://doi.org/10.1371/journal.pmed.1001015.

Togo, A.*et al.* (2019) 'Repertoire of human breast and milk microbiota: a systematic review', *Future microbiology*, 14, pp. 623-641. Available at: https://doi:10.2217/fmb-2018- 0317

Tolossa, T. *et al.* (2020) 'Magnitude and factors associated with lost to follow-up among women under option B+ PMTCT program at East Wollega public health facilities, western Ethiopia', *International Journal of Africa Nursing Sciences*, 13:100212. Available at: https://doi.org/10.1016/j.ijans.2020.100212.

Toniolo, A. *et al.* (1995) 'Productive HIV-1 infection of normal human mammary epithelial cells', *AIDS*, 9(8), pp. 859–866. Available at: https://doi.org/10.1097/00002030-199508000-00005.

Trypsteen, W. *et al.* (2016) 'Diagnostic utility of droplet digital PCR for HIV reservoir quantification', *Journal of Virus Eradication*, 2(3), pp. 162–169. Available at: https://doi.org/10.1016/S2055-6640(20)30460-X.

Tuaillon, E., Al Tabaa, Y., *et al.* (2009) 'Close association of CD8+/CD38bright with HIV-1 replication and complex relationship with CD4+ T-cell count', *Cytometry Part B: Clinical Cytometry*, 76B(4), pp. 249–260. Available at: https://doi.org/10.1002/cyto.b.20467.

Tuaillon, E., Valea, D., *et al.* (2009) 'Human Milk-Derived B Cells: A Highly Activated Switched Memory Cell Population Primed to Secrete Antibodies', *The Journal of Immunology*, 182(11), pp. 7155–7162. Available at: https://doi.org/10.4049/jimmunol.0803107.

Valea, D. *et al.* (2011) 'CD4+T cells spontaneously producing human immunodeficiency virus type I in breast milk from women with or without antiretroviral drugs', *Retrovirology*, 8(1): 34. Available at: https://doi.org/10.1186/1742-4690-8-34.

Van De Perre, P. *et al.* (2012) 'HIV-1 Reservoirs in Breast Milk and Challenges to Elimination of Breast-Feeding Transmission of HIV-1', *Science Translational Medicine*, 4(143). Available at: https://doi.org/10.1126/scitranslmed.3003327.

Van Loggerenberg, F. *et al.* (2008) 'Establishing a Cohort at High Risk of HIV Infection in South Africa: Challenges and Experiences of the CAPRISA 002 Acute Infection Study', *PLoS ONE*, 3(4): e1954. Available at: https://doi.org/10.1371/journal.pone.0001954.

van Zyl, G., Bale, M.J. and Kearney, M.F. (2018) 'HIV evolution and diversity in ART-treated patients', *Retrovirology*, 15(14). Available at: https://doi.org/10.1186/s12977-018-0395-4.

Vandegraaff, N. *et al.* (2001) 'Kinetics of Human Immunodeficiency Virus Type 1 (HIV) DNA Integration in Acutely Infected Cells as Determined Using a Novel Assay for Detection of Integrated HIV DNA', *Journal of Virology*, 75(22), pp. 11253–11260. Available at: https://doi.org/10.1128/JVI.75.22.11253-11260.2001.

Waitt, C. *et al.* (2019) 'Safety and pharmacokinetics of Dolutegravir in pregnant mothers with HIV infection and their neonates: A randomised trial (DolPHIN-1 study)', *PLOS Medicine*, 16(9):e1002895. Available at: https://doi.org/10.1371/journal.pmed.1002895.

Walmsley, S.L. *et al.* (2013) 'Dolutegravir plus Abacavir–Lamivudine for the Treatment of HIV-1 Infection', *New England Journal of Medicine*, 369(19), pp. 1807–1818. Available at: https://doi.org/10.1056/NEJMoa1215541.

Wei, H. *et al.* (2020) 'Defective HIV-1 envelope gene promotes the evolution of the infectious strain through recombination in vitro', *BMC Infectious Diseases*, 20:569. Available at: https://doi.org/10.1186/s12879-020-05288-w.

Wessels, J. *et al.* (2020) 'The updated South African National Guideline for the Prevention of Mother to Child Transmission of Communicable Infections (2019)', *Southern African Journal of HIV Medicine*, 21(1): a1079. Available at: https://doi.org/10.4102/sajhivmed.v21i1.1079.

Whiley, D.M. and Sloots T.P. (2005) 'Sequence variation in primer targets affects the accuracy of viral quantitative PCR', *Journal of clinical virology,* 34(2), pp. 104-7. Available at: https://doi:10.1016/j.jcv.2005.02.010.

Whitman, M. (2012) *The importance melting temperature in molecular biology applications*. Integrated DNA Technologies (IDT). Available at: https://eu.idtdna.com/pages/education/decoded/article/the-importance-of-tm-in-molecular-biology-applications (Accessed: 21 January 2024).

Whitney, J.B. *et al.* (2014) 'Rapid seeding of the viral reservoir prior to SIV viraemia in rhesus monkeys', *Nature*, 512(7512), pp. 74–77. Available at: https://doi.org/10.1038/nature13594.

Winton, L. M. *et al.* (2002) 'Simultaneous one-tube quantification of host and pathogen DNA with real-time polymerase chain reaction', *Phytopathology,* 92(1), pp. 112-6. Available at: https://doi:10.1094/PHYTO.2002.92.1.112.

Wirt, D.P. *et al.* (1992) 'Activated and memory T lymphocytes in human milk', *Cytometry*, 13(3), pp. 282–290. Available at: https://doi.org/10.1002/cyto.990130310.

World Health Organisation (2016) *Consolidated guidelines on the use of antiretroviral drugs for treating and preventing HIV infection: recommendations for a public health approach – 2nd ed.* Available at: 9789241549684_eng.pdf (who.int) (Accessed on 30 September 2024).

World Health Organisation (2018) *Updated recommendations on first-line and second-line antiretroviral regimens and post-exposure prophylaxis and recommendations on early infant diagnosis of HIV: interim guidance*. Available at: WHO-CDS-HIV-18.18-eng.pdf (Accessed on 05 January 2024).

World Health Organisation (2019) *Update of recommendations on first- and second-line antiretroviral regimens.* Available at: WHO-CDS-HIV-19.15-eng.pdf (Accessed: 20 October 2023).

World Health Organisation (2023) *Infant and young child feeding*. Available at: https://www.who.int/news-room/fact-sheets/detail/infant-and-young-child-feeding (Accessed: 20 June 2024).

Wu, J. *et al.* (2022) 'New insights into the role of ribonuclease P protein subunit p30 from tumor to internal reference', *Frontiers in Oncology*, 12:1018279. Available at: https://doi.org/10.3389/fonc.2022.1018279.

Yang, L., Cambou, M.C. and Nielsen-Saines, K. (2023) 'The End Is in Sight: Current Strategies

for the Elimination of HIV Vertical Transmission', *Current HIV/AIDS reports*, 20(3), pp. 121–130. Available at: https://doi.org/10.1007/s11904-023-00655-z.

Ye, J. *et al.* (2012) 'Primer-BLAST: a tool to design target-specific primers for polymerase chain reaction', *BMC bioinformatics,* 18(13):134. Available at: https://doi:10.1186/1471-2105-13-134.

Zhang, J. *et al.* (2014) 'PEAR: a fast and accurate Illumina Paired-End reAd mergeR', *Bioinformatics*, 30(5), pp. 614–620. Available at: https://doi.org/10.1093/bioinformatics/btt593.