Detection of MegaTAL-induced HIV *pol* Mutation By Droplet Digital PCR

Shu Liang

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2015

Committee:

Keith Jerome

Robert Coombs

Jonathan Tait

Noah Hoffman

Program Authorized to Offer Degree:

Laboratory Medicine

University of Washington

Abstract

Detection of MegaTAL-induced HIV *pol* Mutation By Droplet Digital PCR

Shu Liang

Chair of the Supervisory Committee:
Prof. Keith Jerome
Laboratory Medicine

Reliable detection of low-level mutations in HIV-1 provirus is still a challenge. Current methods regularly used in laboratories, Surveyor assay and bulk sequencing, only have a limit of detection about 10 ~ 20%. In the Jerome lab, a potential cure for HIV, megaTALs, is under development and a method to evaluate its outcome is in need. In this thesis, a droplet digital PCR assay that can detect low-level target mutations in HIV provirus induced by megaTALs was developed. This assay was validated with both plasmid and cell line samples. This assay had demonstrated high analytical linearity ($r^2 > 0.99$) and repeatability (CVs < 7%). The Limit of Blank of this assay is 0.56%. The Limit of Detection of this assay is 1.06%. The Limit of Quantitation of this assay is 2.19%. This assay showed higher accuracy than Surveyor assay. The results of the same sample generated by this assay were comparable to those of clonal sequencing and Illumina sequencing. This assay has approved that ddPCR can be use as a reliable mutation detection method in HIV with high precision and accuracy.

Table of Contents

Figure List

21. Surveyor assay steps

Table List

16. Reproducibility and background when detecting mutants from massive WT.
17. Within-run and between-run CVs for Single-Amplicon Assay B
18. Repeatability (CVs) for Single-Amplicon Assay B. (A) Within-run CVs; (B) Between-run CVs.
19. Mutation rate estimated by ddPCR, within-run and between-run CVs for all megaTAL treated samples
20. Mutation rate estimated with Surveyor assay
21. Results form clone sequencing
22. Number of reads for each sample sequenced by Illumina
23. Illumina sequencing results.
24. Mutation affecting REF probe/Fwd Primer/Rev primer binding regions
25. Result comparison from different methods.

Chapter 1: Introduction to Human Immunodeficiency Virus

HIV/AIDS Basics

The human immunodeficiency virus (HIV-1) is a lentivirus, a class of retrovirus that delivers viral RNA into the host's cells causing chronic illness as a consequence. The disease caused by HIV-1 is acquired immunodeficiency syndrome (AIDS), a progressive immune system failure that makes infected individuals susceptible to lethal infections and cancer.[1] About 95% HIV patients live in developing countries, where poverty, social inequalities and poor health care systems are common and facilitate the spread of HIV. HIV/AIDS disrupts community life, interrupts civil order and corrupts economic growth. This epidemic has left millions of orphans, who are in need of basic care and social support. Some of them suffer from HIV infection themselves.[2]

HIV Life Cycle

The major target for HIV is activated CD4 T Lymphocytes. Other cells, like resting CD4 T cells, monocytes and macrophages, and dendritic cells, which also display CD4 plus CCR5 or CXCR4 on the cell surface, can be targeted as well. After interacting with these cell receptors, HIV injects its RNA into the cell, along with several enzymes, including reverse transcriptase, integrase, ribonuclease and protease. [3]

Shortly after HIV entering the target cell, the virus reverse transcriptase uses the HIV genomic RNA creates a "pre-integration" complex in the nucleocapsid, which later travels into the nucleus and integrates into the host genome. This reverse transcription step is highly error prone, which leads to a very high mutation rate in the HIV genome. DNA polymerase then completes the double strand DNA (dsDNA), and then another viral enzyme, integrase, integrates the

dsDNA into the host cell genome, creating the provirus, which is the template for transcription. [4] HIV replicates by releasing full length mRNA comprising the HIV into cytoplasm, where it complexes to structural proteins (Gag, Gag-Pol etc.) and gets packaged into new circular virus particles. After budding from the host cell, proteolytic cleavage of the protein components takes place, and a mature, bullet-shaped capsid HIV is produced. Only mature HIV has the ability to infect other cells.[5]

Pathogenic Dynamics

The progressive depletion of CD4 T cells is the major consequence of HIV infection, which leads to lethal immune dysfunction. This immune system depletion is due to both the death of CD4 T cells and the decrease in CD4 T cell production. Both cell death from HIV infection and bystander effects of syncytia formation, immune activation, proliferation and senescence cause the decrease of CD4 T cells. [6] Moreover, during HIV infection, other cells that facilitate the function of T cells undergo major changes: for instance, there is a profound loss of T-helper-17 (Th17) cells[7] and mucosal-associated invariant T cells, both of which are crucial for defeating bacteria.[8] The immune dysfunction caused by HIV makes patients vulnerable to infections and cancers.

An increase in immune activation is another characteristic of HIV infection. For instance, as a ligand, HIV can trigger the Toll-like receptor 7, leading to the production of interferon (IFN-α).[9, 10]

Treatments

Typically, antiretroviral therapies (ARTs) are used in treating patients with HIV. There are five classes of antiretroviral therapies, targeting different stage of HIV life cycle, that are frequently

used: entry/fusion inhibitors, nucleoside/nucleotide reverse transcriptase inhibitors (NRTIs), non-nucleoside reverse transcriptase inhibitors (NNRTIs), integrase inhibitors and protease inhibitors (PIs). The short-term outcomes of such therapies are usually good. However, patients have to permanently stay on therapy, and there is a possibility of drug resistance due to the high mutation rate of HIV, which can lead to a failure of treatment. Cocktail therapy, a mixture of multiple antiretroviral treatments, is frequently applied to prevent potential drug resistance. Recently, tremendous effort has been put into the discovery of a cure for HIV-- either a functional cure (long-term control without using antiretroviral therapy) or a sterilizing cure (complete elimination of HIV from the patient).

Studies show that early initiation of antiretroviral therapy might preserve immune function and slow establishment of the reservoir.[11,12,13] One approach to HIV cure under study is elimination of latently infected T cells which are present during highly active antiretroviral therapy (HAART).[14] Studies show that the mean half life of the latent reservoir is very long,[15] which is a challenge to achieving a long term cure. Most efforts to eliminate latently HIV infected cells involve deliberately reactivating latent HIV, such that the reactivated cells can become targets for elimination. [16] Another study shows that stimulation of HIV-1-specific cytolytic T cell responses prior to virus reactivation may be essential for successful eradication efforts.[17]

Gene therapy is another avenue towards potential HIV cure. Two possible approaches that have been most widely studied are interrupting the HIV genome or inhibiting HIV replication through host receptor gene disruption.[18] To interrupt the HIV genome, a double strand break (DSB) is created within a desired HIV gene, using any of several classes of highly specific rare-cutting endonucleases. After the DSB is achieved, imprecise repair may introduce a deletion or insertion mutation at the break site. [19,20] Such approach has been applied on herpes simplex virus

(HSV) latent cells in the Jerome lab.[21] A engineered homing endonuclease (HE), which recognizes a sequence in the HSV-1 gene UL19, has showed that HSV production by treated latently infected cells after reactivation was decreased.[21]

Host genes are also popular targets when it comes to HIV gene therapy. For instance, by interrupting HIV receptors such as CCR5 and CXCR4, the entry of HIV can be inhibited, therefore interrupting HIV replication.[22,23]There is a long list of gene-targeting techniques that are feasible for this purpose. Zinc finger nucleases (ZNFs), for example, have been used to target the HIV co-receptor CCR5. After the zinc finger DNA-binding domain has bound to the targeted site, the nucleases generates a DSB in the CCR5 coding region.[24] Transcription activator-like effector nucleases (TALENs), where the TAL effector acts as sequence-specific domain, can be applied in HIV therapy in a similar manner.[25] Large numbers of studies have been done in this field, and gene therapies have already made their way to clinical trails.

Chapter 2: Introduction to droplet digital Polymerase Chain Reaction (ddPCR)

Basic Concepts in Digital PCR

The concept of digital PCR was first described in the early 1990's. [26,26] Similar to traditional quantitative PCR (qPCR), digital PCR requires primers and probes to detect target sequences. However, digital PCR (dPCR) has the potential for increased sensitivity and precision over qPCR, yielding superior results. The major difference is that before PCR cycling to endpoint, samples are randomly distributed into small partitions (nanoliter-size droplets). After endpoint PCR, the presence or absence of template DNA in each partition is determined individually. Since the partitioning is random the number of molecules in each partition follows a Poisson distribution so that positive/negative results can be translated into absolute quantification of template molecules using Poisson statistics. [27]

Digital PCR Platforms

The major differences in all dPCR platforms lie in how the partitions are formed. The partitioning of nucleic acid molecules can be achieved in micro wells, capillaries, the dispersed phase of an emulsion, arrays of miniaturized chambers, or nucleic acid binding surfaces. There are four different commercially available dPCR platforms. Fluidigm Corporation (San Francisco, CA) and Life Technologies (Carlsbad, CA) have each developed a microfluidics-based system that deliver nanoliters of sample into individual reaction wells on sophisticated chips. However, there are a limited number of partitions the sample can be digitized into. [28] Bio-Rad Laboratories (Hercules, CA) and RainDance (Lexington, MA) have developed systems that

partition samples using water-in-oil emulsion droplets (droplet digital PCR, ddPCR). [29, 30] In general, droplet digital PCR platforms allow higher number of partitions at a relatively lower cost, compared to micro-well based platforms[28], making them a better choice for this study.

Advantages and Disadvantages of ddPCR

Compared to real-time qPCR, dPCR in general provides a simplified way to achieve absolute quantification. Unlike qPCR, dPCR quantifies template DNA molecules without the need for a standard curve, and the digitization steps enables quantitation of nucleic acids in a sample with higher precision and a potentially lower limit of detection.[27] For the same reason, dPCR offers sensitive linear response to the small changes in number of copies, which allows more sensitive and accurate detection of copy number variance than qPCR. dPCR increases signal-to-noise ratio in individual partitions, enabling the detection of rare targets. Also, dPCR shows higher tolerance to PCR inhibitory substances, so certain samples that are challenging for qPCR to analyze can be tested by dPCR.[31]

Compared to other dPCR platforms, ddPCR has its own advantages. The dPCR platforms based on micro-well chips have a relatively small number of partitions so the numbers of molecules that can be quantitated in one reaction is limited. To scale up, multiple consumable chips must be used for one sample, which can be expensive. For ddPCR, the cost to scale up for one sample is significantly lower. Therefore, even though the input of molecules in one reaction is still limited, the total input for one sample can be scaled up easily.

Ideally, commercial TaqMan assays should yield reliable results with dPCR platforms too. However, optimization steps may be required when a qPCR assay is transferred to dPCR platforms, since compatibility of a dPCR master mix with primers and probes must be determined empirically. Also, dPCR platforms add some complexity to performing the assay

compared to qPCR. Longer hands-on time is required for dPCR, which may slow down the workflow of a busy lab. More steps of dilution, pipetting and transferring also introduces increased potential for error and contamination.[32]

Applications of ddPCR

The high precision offered by ddPCR enables a long list of applications. One can directly utilize the absolute quantification to detect target DNA, such as viral load analysis and microbial quantification.[33] ddPCR can detect viral sequences with higher precision than qPCR.[34–36] In a Human cytomegalovirus (CMV) study, ddPCR offered increased precision over qPCR when viral load was higher then 4 $\log_{10}$, with similar sensitivity.[33] Rare mutation detection is more likely achieved by ddPCR. Many studies have shown that ddPCR has great utility in detecting oncogene mutations.[37,38] In a non-small cell lung cancer study, ddPCR was able to detect EGFR mutations from plasma samples with a threshold of 1 copy/ml.[37] ddPCR can also detect abnormal fetal genomes from maternal plasma, which enables non-invasive prenatal diagnosis.[39,40] ddPCR has higher sensitivity when detecting copy number variance. Measurement of a 1.2 fold difference in gene copy number can be achieved by ddPCR, the limit of detection of which is significantly lower than other methods.[41] In viral diagnostics, this advantage can be applied to identify chromosomally integrated human herpesvirus 6 (ciHHV6) from active infections to avoid unnecessary treatment.[42] Other applications include quantification of next-generation sequencing (NGS) libraries, gene expression studies, microRNA analysis and single cell analysis.

Chapter 3: Study Rationale, Significance & Approach

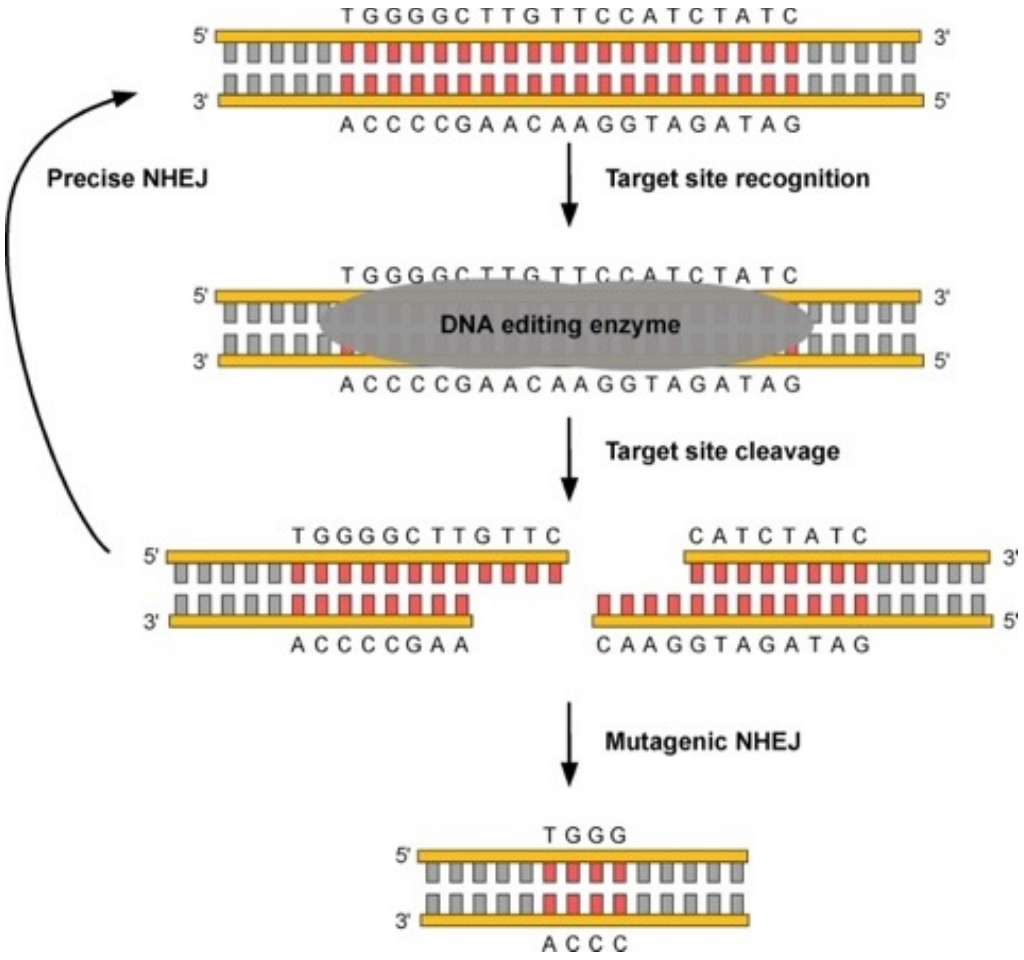
In this study, I evaluated the analytical capability of ddPCR for detecting mutations in HIV provirus. The targeted mutations in this study were those introduced by HIV-targeting megaTALs, one gene therapy method under study in our laboratory for curing HIV. The use of ddPCR may also lead to assays useful for detecting drug resistance mutations in the HIV provirus that are important for selecting the appropriate conventional antiviral treatment.

Target mutations engineered using megaTALs

MegaTALs are a class of newly developed enzymes, first described in 2013.[43] They are a combination of two distinct classes of DNA targeting enzymes, meganucleases and transcription activator-like (TAL) proteins. Meganucleases are a family of naturally occurring rare-cutting endonucleases, characterized by large recognition sites and high cleavage efficiency. The challenge in using meganucleases as a gene therapy is that they are difficult to engineer with high DNA binding efficiency. Therefore, the addition of TAL DNA-binding domains is helpful to increase overall binding affinity.[40,44] TAL effectors are DNA recognizing proteins in which the TAL DNA-binding domains are fused to separate non-specific FokI DNA endonuclease domains in order to achieve a targeted DNA double strand break. The binding domain of TAL effectors can be easily modified to target specific sequences. However, as a gene therapy, TAL effectors require the delivery of two separate proteins into the same cell.[45] Also, strong off-target activities can be observed with TAL effectors. [40] As a fusion of highly-specific meganucleases and the binding domain of TAL effectors, megaTAL allows extremely active and highly specific double strand activity that is applicable to all current viral and non-viral cell delivery methods.[40]

After megaTAL achieves a double strand break, insertions or deletions (indels) at the site of a DNA break are introduced because of mutagenic non-homologous end joining (NHEJ), a pathway that repairs double-strand breaks in DNA. The indels introduced by NHEJ can lead to nonsense and missense mutations in genes to introduce targeted gene disruption.[44,46] A figure describing NHEJ from Schiffer JT is shown below.[46]

In the Jerome lab, the megaTAL under development targets a region in the HIV *pol* gene, which we have designated HIV-S20. S20 is part of the gene that transcribes HIV integrase. By disrupting the formation and function of integrase, the life cycle of HIV is interrupted, leading to a potential cure of HIV infection.



Schiffer JT et. al.,Targeted DNA mutagenesis for the cure of chronic viral infections. 2012

Importance Of A High Precision Diagnostic Method In Patient Management - Drug Resistance

An assay that is able to detect mutations in HIV provirus with high precision can be beneficial in terms of HIV infected patient management. HIV reverse transcriptase has an average error rate per detectable nucleotide of around 1/1700, which is the highest error rate ever described.[47] With such a high mutation rate, HIV is able to escape therapies by developing drug resistance mutations. When using a single time point analysis, a national study in the UK showed that among HIV viremic patients the proportion of results with any type of drug resistance was about 80%[48]. A study conducted in 2004 estimated that the percentage of American HIV positive patient that has any kind of drug resistance related mutation to be 76.3%.[49] Another study showed 56.7% of viremic patients had at least one drug resistance mutation and 37.6% had dual class resistance to nucleoside reverse transcriptase inhibitors (NRTIs) and non-nucleoside reverse transcriptase inhibitors (NNRTI).[50] The possibility of developing drug resistance is significantly higher in patients with previous antiretroviral drug use, advanced HIV diseases, high plasma viral load and low T cell count.[49] Resistance of a therapy can lead to failure of the treatment and progression of AIDS.

When detecting drug resistance related mutations, current Sanger sequencing-based methods have a mutation limit of detection around 20%, meaning only mutations present in 20% or higher of the viruses comprising the total viral load can be detected. Since drug resistance related mutations are accumulating over time, an assay with lower limit of detection allows diagnoses for such mutations sooner, which will benefit patients management significantly. Many studies on drug resistance mutations are underway, and a list of HIV drug resistance mutations is being formed, which is available from an HIV drug resistance database. With such assay, we can

utilize this knowledge better, making meaningful clinical decisions earlier to avoid the use of drug resistance inducing therapies.

Chapter 4 Methods

The assay development was composed with four major steps: assay design, protocol establishment, plasmid sample validation and cellular sample validation. Plasmids and control cell lines with mutations were engineered for assay validation. Assays that passed two validation steps were used to analyzed research samples, the results from which were compared with results from three other mutation detection methods: Surveyor assay, clonal sequencing and Illumina sequencing.

Assay Development using Mutant Plasmid Samples

Mutations Introduced into HIV Sequence

The MegaTAL our laboratory developed with collaborators targets the HIV *pol* gene; the MegaTAL cleavage site was centered on four nucleotides (the central four) around which mutations were introduced. The targeted site sequence is showed in Figure 1. Multiple types of insertion and deletion mutations (indels) induced by the MegaTAL were previously detected by clonal sequencing of MegaTAL-treated integrated HIV genomes (data from Harshana S De Silva Feelixge). Three of these deletion mutations (1bp, 3bp and 7bp deletions) were chosen for further study in this project. Plasmids containing these three deletion mutations as well as a 2bp deletion and a 4bp deletion within the central four were designed to represent likely potential deletion mutations (Fig. 1A).

ddPCR Assay Development: Primers and Probes Design

Three primer and TaqMan MGB probe sets were designed using *Primer Express 3.0* (Life Technologies Inc.). In all three designs, a target probe and a reference (REF) probe were

designed to detect mutations. In each design, the Target probe is complementary to the region containing the megaTAL site; when a deletion or missense mutation occurs the target probe cannot bind the mutant sequence, therefore no signal would be observed. A REF probe that is specific to an HIV genomic region that is not likely to be affected by megaTAL was also included in all three designs, so every HIV sequence should give out REF signal.

Single-Amplicon Assay A	Sequence
Fwd Primer	CAATCCCCAAAGTCAAGGAGTAA
Rev Primer	CCAATCCCCCCTTTTCTTTTA
REF Probe	VIC-AATCTATGAATAAAGAATTAAAG-MGBNFQ
WT Probe	6FAM-TTGTGGATGAATACTGC-MGBNFQ

Double-Amplicon Assay	Sequence
HIV_Ref_F	GTACATACAGACAATGGCAGCAATT
HIV_Ref_R	CCCCGCCCACCAACA
HIV_Ref_VIC	VIC-CCAGTACTACAGTTAAGGC-MGBNFQ
HIV_WT_F	GAGATCAGGCTGAACATCTTAAGACA
HIV_WT_R	CAATCCCCCCTTTTCTTTTAAAA
HIV_WT_FAM	6FAM-TGGCAGTATTCATCCACA-MGBNFQ

Single-Amplicon Assay B	Sequence
Fwd Primer	GGACAGGTAAGAGATCAGGCTGA
Rev Primer	CCAATCCCCCCTTTTCTTTTA
REF Probe	VIC-CATCTTAAGACAGCAGTACAA-MGBNFQ
WT Probe	6FAM-TTGTGGATGAATACTGC-MGBNFQ

Table 1| Primers and probes sequences for 3 ddPCR assay designs.

Two approaches were initially taken, differing in primer design. For Single-Amplicon Assay A and Single-Amplicon Assay B (Ruth H Sedlak), two primers were designed in the same amplicon, using both DNA strands as templates. The difference between these two designs lies in

the distance of the primers and probes. However, in Double-Amplicon Assay the two probes bind to two amplicons separately. The relative positioning of the primers and probes is shown in Figure.2. The primers and probes are listed in Table 1.

All primers were resuspended with 10mM Tris pH 8 into 332 uM primer stocks.

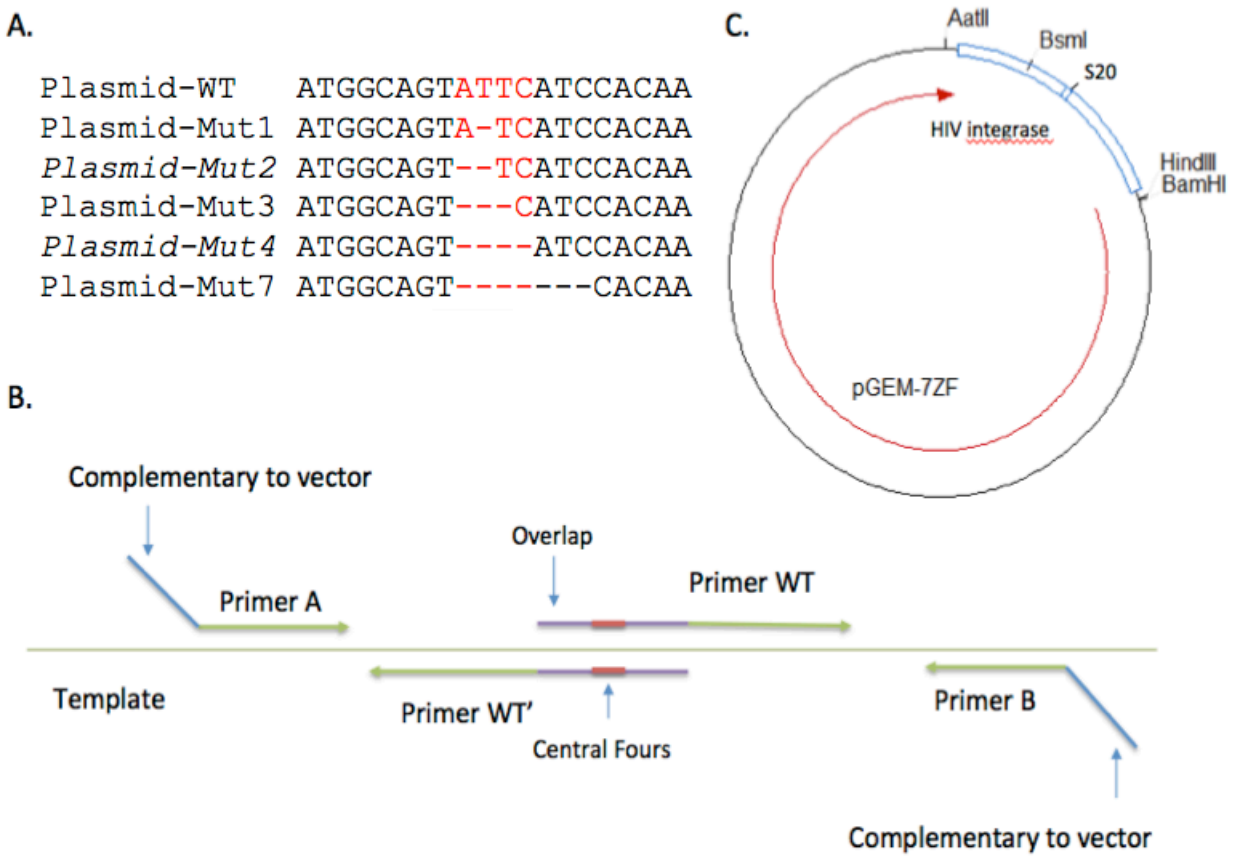
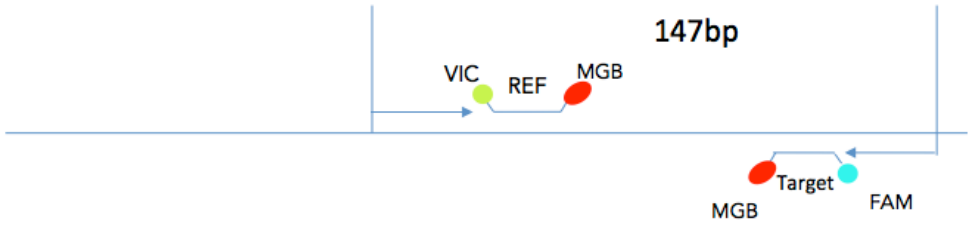
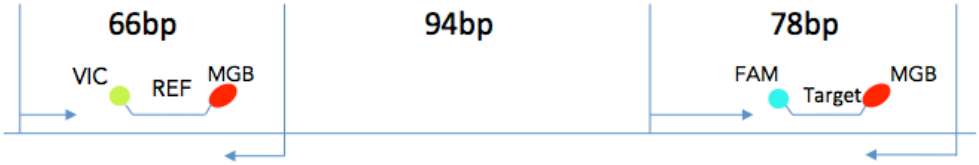


Figure 1 | Plasmid design (A) The mutations that are introduced in HIV *pol* by megaTAL treatment. The position of the central four is in red. (B) Design of primers to amplify fragments for Gibson Assembly from wild type HIV sequence template. This figure shows the specific design for assembly of wild type plasmid (Plasmid_WT). There are two universal primers (Primer A and Primer B) that each have a complimentary part for the template (shown in green) that function as primers in PCR and a complementary part for pGEM-7ZF (shown in blue) that allow assembly of the fragments and vector. Primer WT and Primer WT' have an overlap where the central four (shown in red) were centered. This overlap allows introduction of mutations and the assembly of these two fragments. The lengths of the two fragments are 339bp and 335bp. The total insertion has a length of 646bp. (C) Plasmid map of Plasmid_WT. Sites of restrict enzymes that were used in this project are shown in the map.

A.



B.



C.

Figure 2 | Positioning of primers and probes. Primers are shown as arrows, Target probes are marked as "Target" and REF probes are marked as "REF". For Single-Amplicon Assay A, there is a 147bp amplicon and the two probes are 57bp apart. In Double-Amplicon Assay design, the amplicon that REF probe will bind to is 66bp long, while the amplicon that contains the WT probe binding site is 78bp long. The two regions are 120bp apart. Single Amplicon Design B is very similar to the other single amplicon assay, just with a shorter amplicon length. With all three designs, wild type HIV sequences will present double positive signals, while mutant sequences will only present REF positive signal. With such assay design all REF single positive droplets are identified containing mutant sequences.

Gibson Assembly

Mutant plasmid samples were used in the early phase of assay development. These plasmids were obtained via Gibson assembly, a technique that allows for assembly of multiple DNA fragments, as shown in Figure 3.[51]

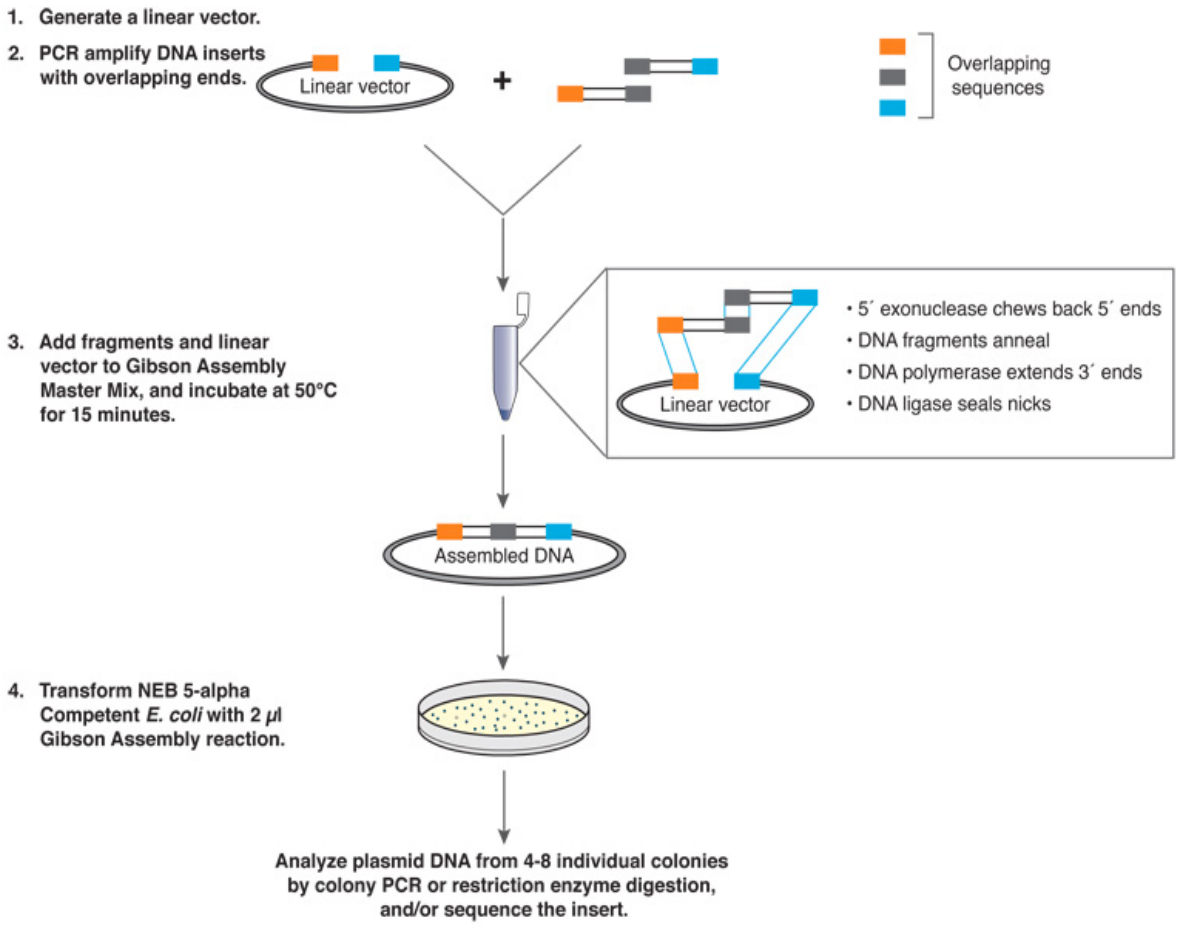


Figure 3 | Major steps of Gibson assembly cloning. Multiple DNA fragments with overlapping ends, one of which can be a linearized vector, can be linked in one reaction. There are three enzymes acting together in Gibson assembly: 5' exonuclease, DNA polymerase and DNA ligase. The 5' exonuclease chews back the 5' end and exposes the complementary sequence in the overlapping ends for annealing. The polymerase activity fills the gaps after annealing and DNA

ligase seals the nick and links the fragments together. After cloning steps, plasmid containing the fragments sequences is obtained.

Image from: http://www.neb.uk.com/Product_Overview/Gibson_Assembly.asp

Fragments from PCR

The fragments to be inserted into the vector are PCR products from a wild type HIV sequence template with primers (Table 2) into which we have introduced specific designed mutations (Fig. 3B). A high-fidelity DNA polymerase (Phusion, Thermo Fisher Scientific.Inc.) was used to minimize the chance of introducing random mutations during PCR. PCR set-up and program are shown in Table 3. All products were confirmed by 2.5% agarose gel electrophoresis at 120V. Gel extraction was performed with Qiaquick Gel Extraction Kit (QIAGEN) to a final volume of 50ul. The concentration of DNA was measured with a NanoDrop 2000 (Thermo Fisher Scientific Inc.). The concentration of all fragments was in 4-8 ng/ul range.

Vector preparation

pGEM-7ZF (2997bp) was used as the vector. 5ug pGEM-7ZF was digested by KpnI overnight in a volume of 50ul. Afterwards the linearized vector was purified by 1% agarose gel electrophoresis at 120V and gel extraction was performed with QIAquick Gel Extraction Kit (QIAGEN) to a final volume of 60ul. The concentration of DNA was determined with a NanoDrop 2000 (Thermo Fisher Scientific Inc.). The concentration of the vector was around 47.3 ng/ul

Primer Name	Sequence
Primer WT	AATGGCAGTATTCATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer WT'	TAAAATTGTGGATGAATACTGCCATTTGTACTGCTGTCTTAAGATG
Primer 1	AATGGCAGTAT-CATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer 1'	TAAAATTGTGGATG-ATACTGCCATTTGTACTGCTGTCTTAAGATG
Primer 2	AATGGCAGT--TCATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer 2'	TAAAATTGTGGATGA--ACTGCCATTTGTACTGCTGTCTTAAGATG
Primer 3	AATGGCAG---TCATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer 3'	TAAAATTGTGGATGA---CTGCCATTTGTACTGCTGTCTTAAGATG
Primer 4	AATGGCAGT----ATCCACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer 4'	TAAAATTGTGGAT----ACTGCCATTTGTACTGCTGTCTTAAGATG
Primer 7	AATGGCAGT-------CACAATTTTAAAAGAAAAGGGGGGATTGGG
Primer 7'	TAAAATTGTG--------ACTGCCATTTGTACTGCTGTCTTAAGATG
Primer A	CTCTAGACTCGAGGAATTCG*CTTGGTAGCAGTTCATGTAG*
Primer B	TTATCGATTTCGAACCCGGG*AATCCTCATCCTGTCTACTT*

Table 2 | Primers for Gibson assembly fragment preparation. The central four and mutations are highlighted in red. The sequences underlined are the overlapping part of the two fragments. The sequences in blue are the complementary sequences for pGEM-7ZF

Assembly and Cloning

Gibson Assembly Cloning Kit (NEB) was used. Gibson Assembly Protocol (E5510) and Gibson Assembly Transformation Protocol (E5510) were followed. 50ng (1.05ul) pGEM-7ZF was used for each assembly reaction. About three fold more insertion fragment molecules were added into the reaction. Total reaction volume was 20ul. Volume of insertion fragments was calculated accordingly.

Screening and Sequencing

Four colonies were selected from each Ampicillin plate and incubated in 2ml Luria Broth (LB) Amp overnight. Plasmids were extracted with PureLink Quick Plasmid Miniprep Kit (Life Technologies). The manufacturer recommended protocol was followed except plasmids were eluted into a final volume of 50ul with TE buffer.

All plasmid samples were screened by restriction enzyme digestion with AatII and BamHI (Fig. 3C), and agarose gel electrophoresis was performed to insure two fragments with length around 2939 bp and 701bp are visible.

After screening, all samples were sequenced using Sanger sequencing. For each mutant plasmid, there was at least one colony containing the intended sequence.

Plasmid Samples Purification

Clones containing the correct sequence were incubated on a larger scale (50ml LB Amp) overnight. Plasmids were extracted with PureLink Quick Plasmid Midiprep Kit (Life Technologies). Manufacturer recommended protocol was followed except the two centrifuge steps were changed into 30min at 4000rpm and 45min at 4000rpm.

Thermocycling conditions		
98°C	30 sec	
98°C	10 sec	30 cycles
60°C	15 sec	
72°C	15 sec	
72°C	5min	
4°C	Hold	

PCR reaction set up	
10uM Primer*	2.5ul
10uM Primer A or Primer B	2.5ul
5X Buffer	10ul
10Mm dNTPs	1ul
Enzyme, Phusion	0.5ul
Template	1ul (10ng)
Water	32.5ul
Total	50ul

Table 3 | PCR program and reaction set up for Gibson assembly fragment preparation. "Primer*" stands for the primer used with Primer A or Primer B to amplify a particular fragments. For example, to get the two fragments used in Gibson assembly to form Plasmid_WT, two PCRs were set up. Primer WT' and Primer A was used to amplify one fragment, and Primer WT and Primer B were used to amplify the other fragment.

Plasmid Sample Pre-analysis Preparation

Plasmid samples were linearized with a single cutter, HindIII (New England Biolabs) (shown in Fig 3.C), before analysis by ddPCR, according to manufacturer's recommendations for plasmid analysis. They were digested for 2 hours at 36°C. A 20ul reaction contained 17.5 ul DNA, 0.5 ul HindIII and 2ul 10X Cutsmart buffer.

PCR Condition Optimization

Plasmid_WT was used in all experiments to optimize the PCR conditions for the TaqMan assays. All experiments were performed on QX100 Droplet Digital PCR System (Bio-Rad. Inc.). All accessories, reagents and consumable were from Bio-Rad also.

Reactions with Single-Amplicon Assay A and Double-Amplicon Assay were set up using the Bio-Rad ddPCR Supermix. A gradient PCR was performed with each assay using 5ul of 0.000001ng/ul Plasmid_WT to optimize PCR cycling conditions.

For Single-Amplicon Assay A, a range of 57°C to 63°C for annealing temperature was applied but the reaction efficiency of the REF probe was insufficient under all conditions tested.

For Double-Amplicon Assay, a range of 51°C to 63°C for annealing temperature were applied. Both WT probe and REF probe showed optimal efficiency at 52°C, so primers and probes from design B were selected as the report system for this assay. The finalized PCR program is shown in Table 4 and Table 5. The same themocycling conditions and reaction set-up was used throughout further experiments for Double Amplicon Assay.

For Single-Amplicon Assay B, reactions were set up using the Bio-Rad ddPCR Supermix (no dUTP). A gradient PCR was performed using 5ul of 0.000001ng/ul Plasmid_WT to optimize PCR cycling conditions. A range of 48°C to 56°C for annealing temperature was applied. Later,

another gradient PCR covering a range of 53°C to 58°C was performed, with 5ul of a mixture of Plasmid_WT and Plasmid_Mut1 as template, which had a concentration around 0.000001ng/ul.

20X primer-probe mix	
HIV_Ref_F (332uM)	10.84 ul
HIV_Ref_R (332uM)	10.84 ul
HIV_WT_F (332uM)	10.84 ul
HIV_WT_R (332uM)	10.84 ul
HIV_WT_FAM (100uM)	10 ul
HIV_Ref_VIC (100uM)	10 ul
Tris 10mM	136.64 ul
Total	200 ul

PCR reaction set up	
20X primer-probe mix	1.25ul
2X ddPCR master mix	12.5ul
DNA	x ul
Water	(11.25 – x) ul
Total	25ul

Table 4 | Prime-probe mix preparation and reaction set up for Double-Amplicon Assay and Single-Amplicon Assay B.

Thermocycling condition for Double-Amplicon Assay		
95°C	10 min	
94°C	30 sec	40 cycles
52°C	1 min	
98°C	10 min	
4°C	hold	

Thermocycling condition for Single-Amplicon Assay B		
95°C	10 min	
94°C	30 sec	40 cycles
57°C	1 min	
98°C	10 min	
4°C	hold	

Table 5 | Thermocycling conditions for all ddPCR experiments in this project.

Detecting Low Level Mutants from A WT Background (Linearity and Repeatability)

To validate the ability of detecting low-level mutant sequences from a massive wild type background for the ddPCR assays, 2-fold low-level dilution series of mutant plasmid samples (Plasmid_Mut2 or Plasmid_Mut7) were mixed with a constant Plasmid_WT sample. For Single-

Amplicon Assay A and Double-Amplicon Assay, the Plasmid_WT background level was around 0.00001ng/ul, (about 200-270 copy/ul in reaction). For Single-Amplicon Assay B, the Plasmid_WT level was around 490 copy/ul in reaction. Samples were diluted with 10mM Tris pH8.0. On the same ddPCR plate, the dilution series of mutant plasmids alone was analyzed. By comparing the result from Plasmid_WT and Plasmid_Mut Mixture with the result from Plasmid_Mut only wells, linearity analysis was performed. Plasmid_WT samples were analyzed on the same plate as negative controls. All samples were run in triplicate.

Finally, all data from Plasmid_WT only reactions were collected to roughly estimate the level of false mutation rate from known wild type sequences.

Analyzing MegaTAL treated HIV cell line samples by ddPCR

Control Cell Lines Preperation

Three control cell lines with HIV *pol* gene megaTal recognition site were made to validate the assays, HIV megaTAL site WT/Mut1/Mut7. The fragments of interest were inserted into a lentivirus backbone vector, pRRL-MND-GFP, using Gibson Cloning , which later was used to construct lentiviruses. Later, the lentiviruses were used to infect SupT1 cells to make HIV-megaTAL site integrated cell lines.

Lentivirus Backbone Vector Made with Gibson Cloning

The same Gibson Cloning protocol was followed when making lentivirus backbone vectors containing the fragments of interests.

The fragments to be inserted into the lentivirus vector were amplified from the plasmid samples obtained previously, Plasmid_WT, Plasmid_Mut1 and Plasmid Mut7. The primers used are

listed in Table 6. The fragments amplified are the same HIV fragments that were used to construct plasmid samples.

5ug pRRL-MND-GFP was linearized by BstxI overnight in a volume of 50ul. Same purifying steps were performed as previously described. The ddPCR target sequence was inserted between the HIV cPPT sequence and the MND promoter driving expression of GFP.

Primer Name	Sequence
Lenti-polF	GCTTGATATCGAATTCCCACCTTGGTAGCAGTTCATGTAG
Lenti-polR	CTCTGTTCCTACGCGTCCAAAATCCTCATCCTGTCTACTT

Table 6 | | Primers for Gibson assembly fragment preparation. The sequences in blue are the complementary sequences for pRRL-MND-GFP.

Mutant Control Cell Lines

SupT1 cells (ATCC# CRL-1942) were grown in RPMI 1640 (Invitrogen) supplemented with 10% FBS. 293T (ATCC# CRL-3216) cells were grown in DMEM (Invitrogen) supplemented with 10% FBS.

To generate a VSV-G pseudotyped stock of each lentiviral vector, a 10cm dish containing HEK293T cells at 70% confluence was transfected with a total of 10mg of lentivirus vector plasmid and the packaging plasmids psPAX2 and pMD2G. At 72 hours post transfection, cell supernatants were collected and stored at -80^{o}C. Control SupT1 cells with integrated copies of wild type or mutant ddPCR target sites were then generated by transduction with each lentivirus vector. To ensure that the provirus copy number remained below 1 copy/cell, lentivirus transduction levels of less than 30% were chosen.[52] SupT1 cells were infected with equalized

GFP expressing units of each vector, expanded for five days, then sorted for GFP positive cells before genomic DNA extraction using the Qiaqen DNeasy blood and tissue DNA kit. Genomic DNA was extracted from bulk infections so that ddPCR would be performed on control samples with a polyclonal integration site profile.

Cell line preparation was performed by other lab member (Harshana S. De Silva Feeilxge).

Research Samples

MegaTALs with either 6.5 or 7.5 repeat variable diresidue (RVD) repeats, were used to treat cell lines containing either integrated HIV or transfected with HIV plasmid sequence in previous experiments performed by other lab members (Harshana S. De Silva Feeilxge and Nixon Niyonzima). Two of the samples were also treated with Trex2, a three prime repair exonuclease 2, which increases the likelihood of target mutagenesis. (See in Table 7)

gDNA Sample Pre-analysis Preparation

Digestion is recommended by the manufacturer to reduce viscosity of gDNA samples, so packaging of DNA into droplets is improved. HindIII was used to digest some gDNA samples that will be analyzed with both single-amplicon assays. The digestion takes 4 to 6 hours at 36°C. The digestion reactions were set-up in either 20ul or 50ul volume with 10X Cutsmart buffer (New England Biolabs).

In the double-amplicon assay, to unlink the two amplicons, a restriction enzyme, BsmI (New England Biolabs), was used to digest samples before analysis. The BsmI cuts a site in between the two regions that will be amplified, but does not interfere with the sequences within the amplicons. The restriction site for BsmI in Plasmid_WT is shown in Fig 1.C. The new assay design is shown in Figure 4. Samples were digested by BsmI for 4 hours at 65°C.

Digestion Comparison

To confirm that BsmI unlinks the two regions, the sequence between HIV_Ref_F and HIV_WT_R was amplified by these two primers. The thermocycling condition with Phusion was described previously. The PCR product was separated into three tubes and digested with water, HindIII or BsmI for 1 hour. Undigested PCR product, HindIII digested PCR product and BsmI digested PCR product were applied to electrophoresis on the same 1% gel. The gel image was captured afterwards.

Figure 4 | Final Double-Amplicon assay design. DNA is digested by BsmI before analysis. Two regions are randomized separately into droplets and amplified.

Analysis of Nonspecific Amplification

HindIII digested SupT1 was used as negative control to eliminate the possibility that the assay amplifies nonspecific human genome sequences. 1ul SupT1 (34 ng/ul) was applied in each reaction to confirm no nonspecific amplification was observed. Samples were analyzed in triplicates.

Experiments Concerning Formation of Single Positive Droplets

To determine if non-specific amplifications in human genome were the reason of formation of single positive droplets, gDNA from SupT1 cell line was analyzed by Double-Amplicon Assay.

To eliminate the possibility of inhibition from gDNA background to WT and REF probes, a mixture of a two-fold dilution series of HindIII digested Plasmid_WT and a 1ul of HindIII digested SupT1 (34 ng/ul) background and the same Plasmid_WT dilution series were analyzed on the same plate, and the results were compared to determine inhibition for either primer/probe. Samples were analyzed in triplicate.

Cell Line Samples	Description
SupT1-MegaTAL site WT	gDNA from a SupT1 cell line that contains a HIV sequence fragment with wild type megaTAL site
SupT1-MegaTAL site Mut1	gDNA from a SupT1 cell line that contains a HIV sequence fragment with megaTAL site that has 1bp deletion
SupT1-MegaTAL site Mut7	gDNA from a SupT1 cell line that contains a HIV sequence fragment with megaTAL site that has 7bp deletion

Control Sample	Description
SupT1	gDNA from wild type SupT1 cell line
SupT1-DHIV3	gDNA from wild type HIV integrated cell line
293T + DHIV3	Plasmid DNA that contains HIV sequences transfected into 293T cells

MegaTal Treated Sample	Description
SupT1-DHIV3 + 6.5	gDNA from HIV integrated SupT1 cells that are treated with 6.5
pDHIV3 + 6.5	Plasmid DNA that contains HIV sequences transfected into 293T cells and treated with megaTAL 6.5
pDHIV3 + 6.5 + Trex2	Plasmid DNA that contains HIV sequences transfected into 293T cells and treated with megaTAL 6.5 and Trex2
pDHIV3 + 7.5	Plasmid DNA that contains HIV sequences transfected into 293T cells and treated with megaTAL 7.5
pDHIV3 + 7.5 + Trex2	Plasmid DNA that contains HIV sequences transfected into 293T cells and treated with megaTAL 7.5 and Trex2

Table 7 | Plasmid and cell lines cellular samples used in assay validation and analyzed with ddPCR assays

Limit of Blank (LoB) Estimation

LoB was estimated with 60 measurements under the guidelines of established protocol (E-17, NCCLS).

Double-Amplicon Assay

SupT1-DHIV3 was diluted into a six level dilution series that covers the range of 1.6 – 9.6 ng/ul. The dilution factors and correspondent concentration is shown in Table 8. Samples were diluted with 10mM Tris pH8. 10 ul of diluted SupT1-DHIV3 was applied in each reaction. Five identical samples of each dilution level were analyzed on the same plate to address aliquoting variability. Another five identical samples of each dilution level were analyzed separately to address between-run variability.

Dilution Factor	0.1667	0.3333	0.5	0.6667	0.8333	1
Concentration	1.6 ng/ul	3.2 ng/ul	4.8ng/ul	6.2 ng/ul	7.8 ng/ul	9.6 ng/ul

Table 8 | Dilution series of SupT1-DHIV3 used to establish LoB

Single-Amplicon Assay B

SupT1-MegaTAL site WT was diluted into a 5 level dilution series that covers the range of 3.36 – 16.8 ng/ul. The dilution factors and correspondent concentration is shown in Table 9. Samples were diluted with 10mM Tris pH8. 10 ul of diluted SupT1-MegaTAL site WT was applied in each reaction. 4 repeats of each dilution level were analyzed on the same plate to address aliquoting variability.

Three plates with same experiment setup were analyzed separately to address between-run variability.

Dilution Factor	1	0.8	0.6	0.4	0.2
Concentration	16.8ng/ul	13.44ng/ul	10.08ng/ul	6.72ng/ul	3.36ng/ul

Table 9 | Dilution series of SupT1-MegaTAL Site WT used to establish Limit of Blank.

Limit of Detection (LoD) Estimation For Single-Amplicon Assay B

To establish LoD, SupT1-MegaTAL site Mut1 and SupT1-MegaTAL site Mut7 were mixed with SupT1-MegaTAL site WT to get 6 mixtures with different levels of mutation rate, so the range of mutation rate evaluated covers the LoB to five times of LoB. 6 repeats of each mixture were analyzed on one plate. Two identical plates were analyzed on different days. Samples were diluted with 10mM Tris pH 8.

Limit of Quantitation (LoQ) Estimation For Single-Amplicon Assay B

The same experiment set up as LoD was performed to collect more data for LoQ estimation.

Mutation Detection in MegaTAL treated Samples

Before analysis, samples were digested by BsmI for Double-Amplicon assay and by HindIII for Single-Amplicon assay B. The digestion products were quantified by a NanoDrop 2000 (Thermo) then diluted with 10mM Tris pH 8. After dilution, the predicted output of REF probe for each sample would fall in the range of 100 copy/ul to 400 copy/ul. The dilution chart is shown in Table 10. 10 ul of diluted samples were added in each reaction. All samples were

analyzed in triplicates by the method established. See Table 4 for reaction set-up and thermocycling conditions. The analysis was repeated twice, so for each sample, nine measurements were taken.

Sample Name	Original Concentration (ng/ul)	Dilution Factor
SupT1-DHIV3	110.9	0.05
293T + pDHIV3	141.6	0.00025
SupT1-DHIV3 + 6.5	107.5	0.05
pDHIV3 + 6.5	78.2	0.0002
pDHIV3 + 6.5 + Trex2	90.2	0.0004
pDHIV3 + 7.5	84.0	0.00015
pDHIV3 + 7.5 + Trex2	75.4	0.00025

Table 10 | Dilution chart for control and research samples.

Method Comparison With Surveyor Assay, Clonal sequencing and Illumina Sequencing

Surveyor Assay

PCR amplicons spanning the HIV-pol megaTAL target site were generated using Phusion polymerase (New England Biolabs), and primers HIVintF (TAGCAGGAAGATGGCCAGTA) and HIVintR (TCCTGTATGCAGACCCCAAT). PCR products were column purified before 200-400ng of DNA was heteroduplexed by heating to 98^{o}C for ten minutes and then slowly cooling to room temperature before placing on ice. Digestions were performed in a volume of

15 15µl with 5U of Surveyor enzyme at 37°C for 30 minutes before cleavage was analyzed on a 2% agarose gel. Quantification of cleavage was performed using ImageJ software as previously described.[53] Collaborating members of the Jerome lab (Nixon Niyonzima and Harshana S De Silva Feelixge) performed this analysis.

Clonal Sequencing

Clonal amplicon sequencing was performed on DNA extracted from treated cells as previously described.[21,53] ZFN target sites were amplified using Phusion polymerase (New England Biolabs), and primers HIVintF (TAGCAGGAAGATGGCCAGTA) and HIVintR (TCCTGTATGCAGACCCCAAT). PCR products were sub-cloned using the Zero Blunt TOPO PCR cloning kit (Invitrogen). TOPO-cloned PCR products were transformed into One Shot Top10 *Escherichia coli* (Invitrogen) for clonal analysis and individual colonies were picked for plasmid purification from which the clonal ZFN target sites were sequenced using T7 or SP6 sequencing primers. Members of the Jerome lab (Nixon Niyonzima, Harshana S De Silva Feelixge) performed this analysis.

Illumina Sequencing

For library preparation for research samples, a pair of amplicon primers with adapter sequences of 16S ribosomal RNA gene (16S rRNA) was designed (see Table 11). The HIV sequence that was amplified was 327bp. The amplicon primers were tested on SupT1-DHIV. Kappa HiFi polymerase was used to reduce PCR errors. To optimize the PCR reaction, gradient PCR was performed with an annealing temperature range of 60°C to 75°C. An annealing temperature of 61°C showed the best efficiency. The final program and reaction setup is shown in Table 12. SupT1 control sample and all MegaTAL treated samples were amplified with this PCR

condition. NGS library preparation and Illumina sequencing were performed by Roger Bumgarner's laboratory. The sequencing was performed with the Illumina MiSeq system according to the manufacturer's recommended protocol.

For data analysis, DNA sequences were pre-processed using the Geneious software package version R6.18 (www.geneious.com). Paired reads were trimmed to eliminate poor quality sequences at the 3' and 5' ends, and reads shorter than 200 bp in length or with quality scores below 30 were removed. Trimmed and filtered reads were mapped to the DHIV reference sequence and the alignment was exported for variant analysis performed with the Rsamtools, ShortRead and Biostrings packages in R/Bioconductor (3-5). Aligned reads were scanned for insertions, deletions and SNPs, focusing in particular on regions identified as the endonuclease target region (17 bp), forward and reverse primer binding regions (23 bp and 21 bp respectively) and reference region (21 bp). Quality scores of insertions and SNPs were also recorded to identify poor quality calls, and the percentages of reads containing mutations of each type and in each region were counted.

Primer	Sequence
Illseq_Fwd	TCGTCGGCAGCGTCAGATGTGTATAAGAGACAGACAGTTAAGGCCGCCTGTTG
Illseq_Rev	GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAGTGGATCTCTGCTGTCCCTGT

Table 11 | Amplicon primers for Illumina sequencing library preparation. The adapter sequences are underlined.

PCR reaction set up	
10uM Illseq_Fwd	1.5ul
10uM Illseq_Rev	1.5ul
5X KAPA HiFi Buffer	5ul
10Mm dNTPs	1ul
KAPA HiFi	0.5ul
Template	1ul (10ng)
Water	14.5ul
Total	25ul

Thermocycling condition		
95°C	3 min	
98°C	20 sec	30 cycles
61°C	15 sec	
72°C	30 sec	
72°C	1 min	
4°C	Hold	

Table 12 | Reaction set-up and thermocycling conditions for the first step PCR amplification for library preparation.

Chapter 5 Results

Two ddPCR assays were developed to detect megaTAL-induced mutations from the HIV *pol* gene. Validation with plasmid and cell lines was performed using both assays. Research samples were analyzed by Double-Amplicon Assay and Single-Amplicon Assay B, the results of which were later compared with results from other methods (Surveyor assay, clonal sequencing and Illumina sequencing).

Plasmids with wild type or the desired mutant HIV sequence integrated were obtained via Gibson cloning and were used to validate ddPCR assays.

Single-Amplicon Assay A

In Single-Amplicon Assay A, the wild-type (WT) probe showed excellent efficiency when the annealing temperature was 58.3°C, while the reference (REF) probe showed no amplification. (Fig 5) After a few failed attempts of optimizing reaction conditions, further effort on developing Single-Amplicon Assay A was determined pointless.

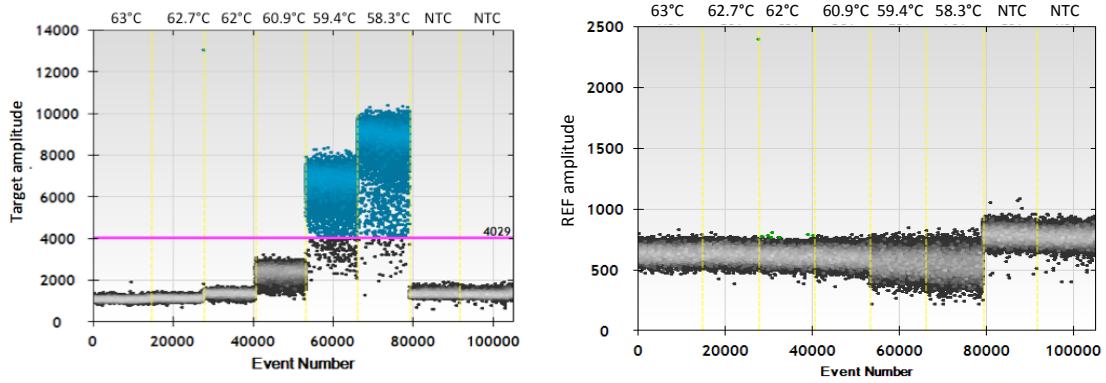


Figure 5 | Gradient PCR results from Single-Amplicon Assay A in 1-D amplitude plot. In Single Amplicon Assay A, REF probe showed no efficiency with any annealing temperature tested.

Double-Amplicon Assay

A double-amplicon assay was developed and validated with both plasmid and cellular samples. Limit of Blank for this assay was estimated.

Assay Development on Mutant Plasmid Samples

Two primer and TaqMan MGB probe sets were designed and tested with the wild-type plasmid (Plasmid-WT). The gradient ddPCR results for both designs are shown in Fig6.

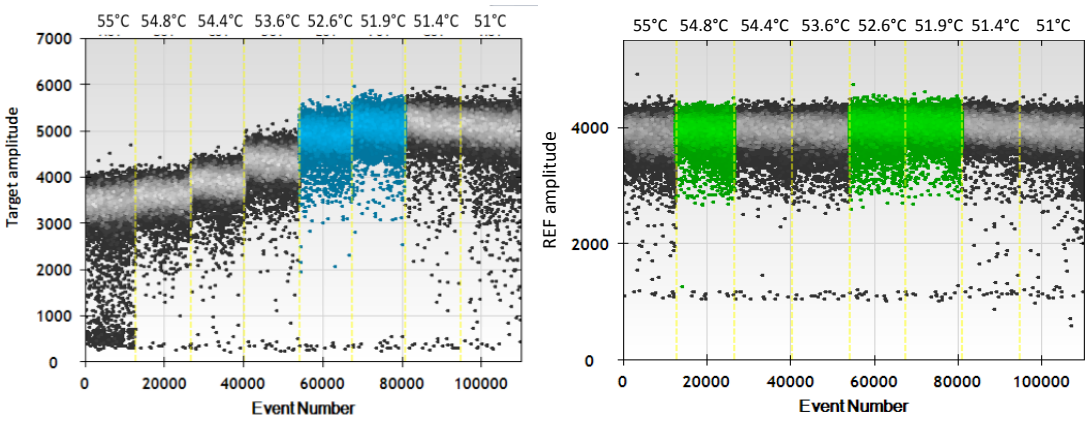


Figure 6 | Gradient PCR results from Double-Amplicon Assay in 1-D amplitude plot. Both Target and REF probe showed the best efficiency and clearest separation between positive and negative droplets with an annealing temperature of 51.9°C.

Double-Amplicon Assay showed optimal efficiency for both the WT and REF probes at 51.9°C. Non-specific amplification was not observed. Therefore, this design proceeded to subsequent validation steps. Different mutant plasmids were observed as different droplet groups, as shown

in Figure 7. With this assay design, all REF single positive droplets were identified as containing mutant sequences (green and orange droplets in Figure 7).

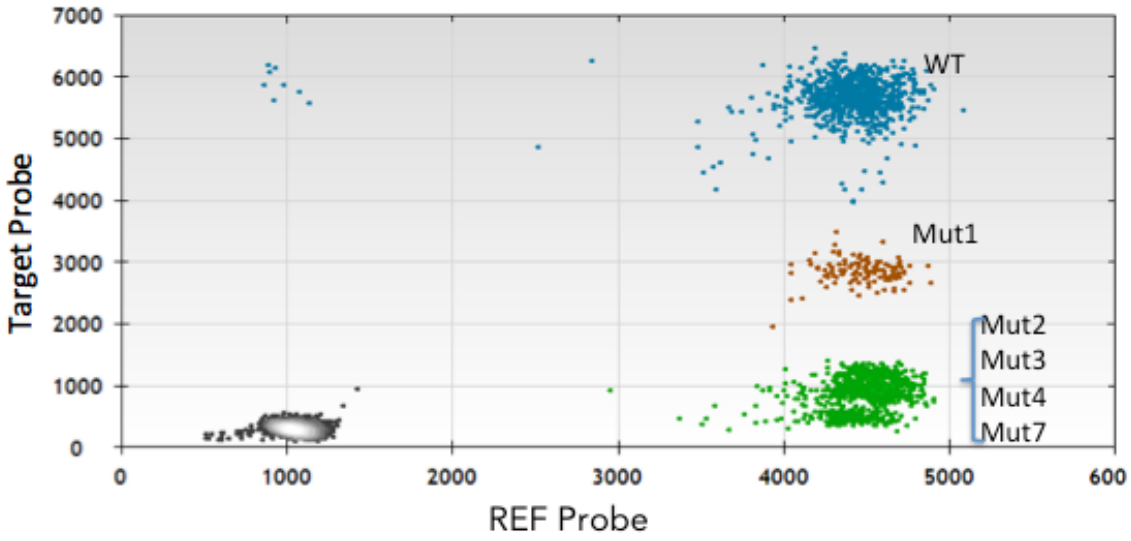


Figure 7 | Different droplet groups observed with plasmid samples in Double-Amplicon Assay. WT: droplet group observed from Plasmid_WT (blue). Mut1: droplet group (orange) observed from Plasmid_Mut1 (with 1bp deletion) presents medium amplitude due to less-efficient binding of the probe to the 1bp deletion sequence. Larger deletions (Mut2-7) abrogate probe binding and exhibit very low amplitude droplets (green).

Detecting Low-Level Mutants from A WT Background (LoB, Linearity and Repeatability)

Plasmid-Mut (Plasmid-Mut2 or Plasmid-Mut7) 2 fold dilution series and constant Plasmid-WT were mixed and analyzed, the results of which were defined as observed mutant sequences. The same Plasmid-Mut series were analyzed simultaneously, the result from which was defined as the input for the mixture. To estimate the capability of detecting low-level mutants, the two sets of results were compared. The estimation of linearity, repeatability (coefficient of variance) and

LoD was performed with the same data sets. The predicted rate of falsely identifying a mutant sequence (background) was estimated with data collected from Plasmid-WT only wells.

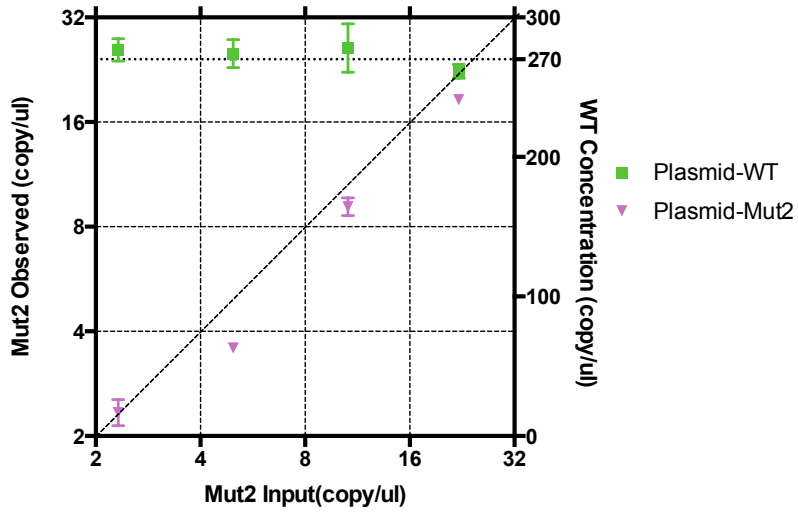
For Plasmid-Mut2, the linearity analysis result had a R square value of 0.9910. The average coefficient of variance (CV) within each triplicate (intra-assay variation) was 6%. The dilution series was mixed with a constant Plasmid-WT level of 270 copy/ul. The predicted rate of falsely identified mutant sequence is less then 0.5% for such Plasmid-WT input. With all this information, a LoD of mutation rate was roughly established at 1% (Fig.8A).

For Plasmid-Mut7, the lowest dilutions (0.564%) gave two readouts as 0 copy/ul, which means the limit of detection was reached. The linearity analysis showed a R square value of 0.9519. The average CV within each triplicate was 8%. The Plasmid-WT level was 200 copy/ul in the mixture, with a predicted background value between 0.5% and 1%. Therefore a LoD of mutation rate was roughly estimated as 1% (Fig.8B).

	Plasmid_WT	R square	intraassay CV	Background	LoD
Mut_2	270 copy/ul	0.9910	6%	0.5%	1%
Mut_7	200 copy/ul	0.9519	8%	0.5 - 1%	1%

Table 13 | Repeatability and limit of detection estimation for Double-Amplicon Assay using plasmid samples.

A.



B.

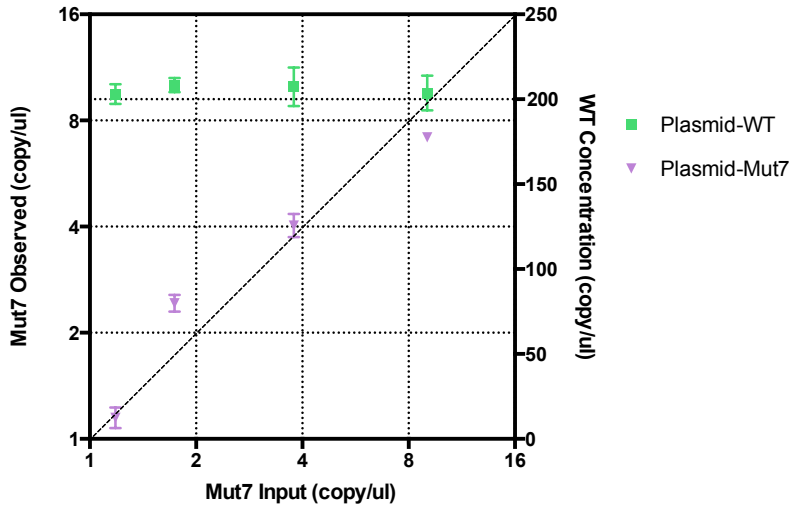


Figure 8 | Linearity, repeatability and limit of detection (LOD) for mutant plasmid samples. Mutant sequences observed from Plasmid-WT and Plasmid-Mut mixture were plotted to left Y-axis, while the results from Plasmid-Mut only wells (mutant sequences input) were plotted to X-axis. The Plasmid-WT level was plotted to the fight T-axis. A. Estimation with Plasmid-Mut2: $r^2 = 0.9910$, CV = 6%, LoD = 1%. B Estimation with Plasmid-Mut2: $r^2 = 0.9519$, CV = 8%, LoD = 1%.

Analyzing cellular samples by ddPCR

As previously described, in the plasmid samples, since there were almost no single positive droplets observed, the read out from the REF probe single positive droplets is defined as mutant. However, when the assay was first applied to a negative control cellular sample (SupT1-DHIV3), which has HIV genome integrated in the cellular gDNA, both WT probe single positive and REF probe single positive droplet groups were observed with a percentage of 4.4%, when compared to the total droplets count, which could interfere the detection of mutant sequences. (Fig.9)

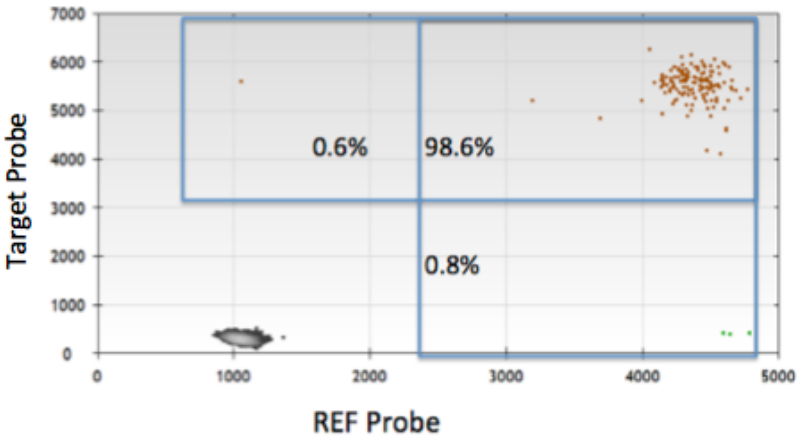
Mechanism of Formation of Single Positive Droplets

We first considered the possibility that single positive droplets could result from nonspecific amplification of human genome sequences. We evaluated this by analyzing SupT1 gDNA with Double-Amplicon Assay. No positive droplets were observed, suggesting that there was no nonspecific amplification. We evaluated two additional potential explanations for single positive droplets: first, probe interference from gDNA which increases the percentage of single positive droplets, or inhibits one probe more significantly then the other; second, fragmentation during cellular DNA extraction.

To verify if inhibition happened in cellular samples, a mixture of a 2-fold dilution series of HindIII digested Plasmid_WT (4 levels) and a constant 1ul of HindIII digested SupT1 (34 ng/ul) background was analyzed while the same Plasmid_WT dilution series was analyzed on the same plate. By comparing the percentage of single positive droplets observed in both experiment set-ups in a two-way ANOVA analysis (single Target positive: $P = 0.2504 > 0.05$, single REF positive: $P = 0.1427 > 0.05$), I demonstrated that gDNA would not increase the possibility of

single positive droplets, thus this is not the reason for the formation of single positive droplets. (Fig.10B)

A.



B.

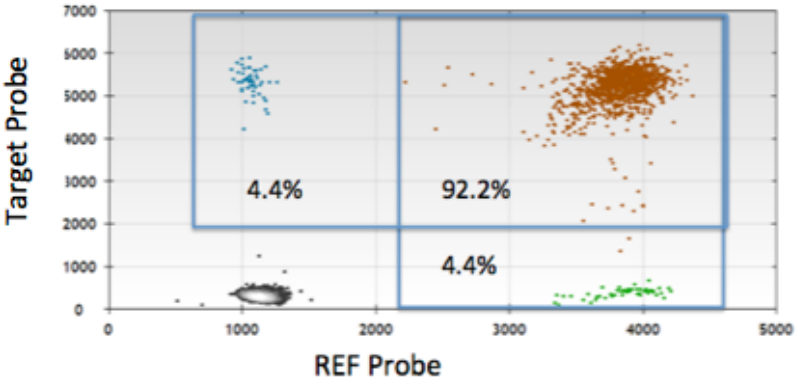


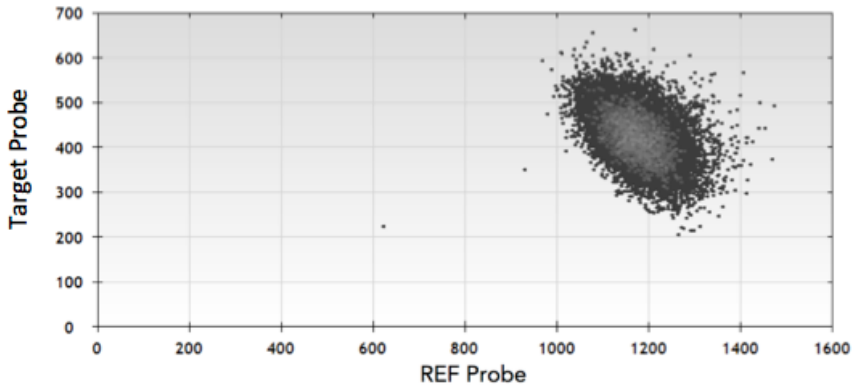
Figure 9 | Single positive droplets were observed in Double-Amplicon Assay. (A) A 2-D amplitude plot from Plasmid_WT. There were few single positive droplets observed but the percentage of which in the total droplet count was low (<1%). (B) A 2-D amplitude plot from SupT1-DHIV3 gDNA sample. Both WT probe and REF probe single positive droplets were observed, with a similar percentage of 4.4%.

Similar inhibition levels (18%) of both probes were observed in every dilution level. Since the inhibition effect is equivalent for both probes ($P = 0.3405 > 0.05$), it is not likely that one probe within a droplet would be inhibited while the other is not, which would present single positive droplets in a reaction. (Fig.10C)

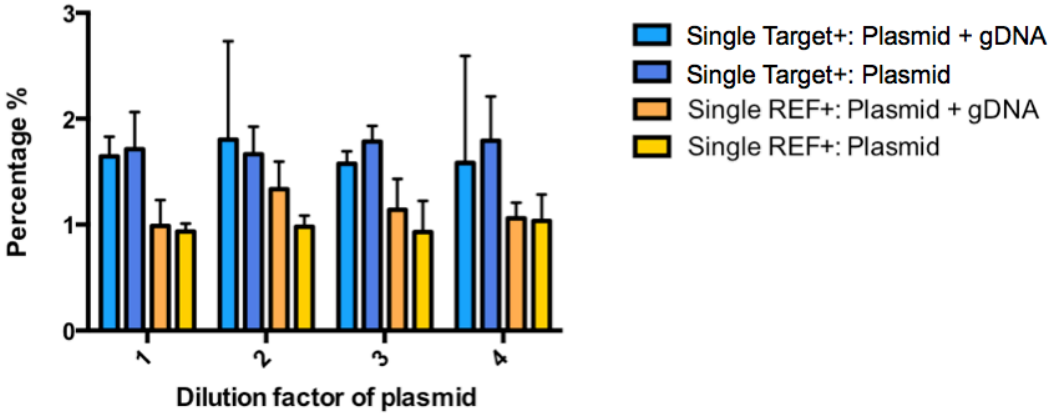
Given that the other hypotheses were eliminated, we concluded that the most likely mechanism for the formation of single positive droplets was DNA fragmentation. While fragmentation of DNA can be evaluated by a number of techniques, one of the most common and simplest techniques is analysis by separate PCR reactions separated by short DNA sequences as performed here. We therefore evaluated techniques to minimize potential fragmentation of DNA during our assay.

To minimize the influence of unexpected DNA fragmentation, a restriction enzyme, BsmI, was used in sample preparation, as shown in Fig.4. The two regions were proved to be separated after BsmI digestion, therefore the signals from two probes can be analyzed separately. (Fig 11)

A.


Target Probe
REF Probe
0
100
200
300
400
500
600
700
0
200
400
600
800
1000
1200
1400
1600

B.


Percentage %
Dilution factor of plasmid
Single Target+: Plasmid + gDNA
Single Target+: Plasmid
Single REF+: Plasmid + gDNA
Single REF+: Plasmid

C.

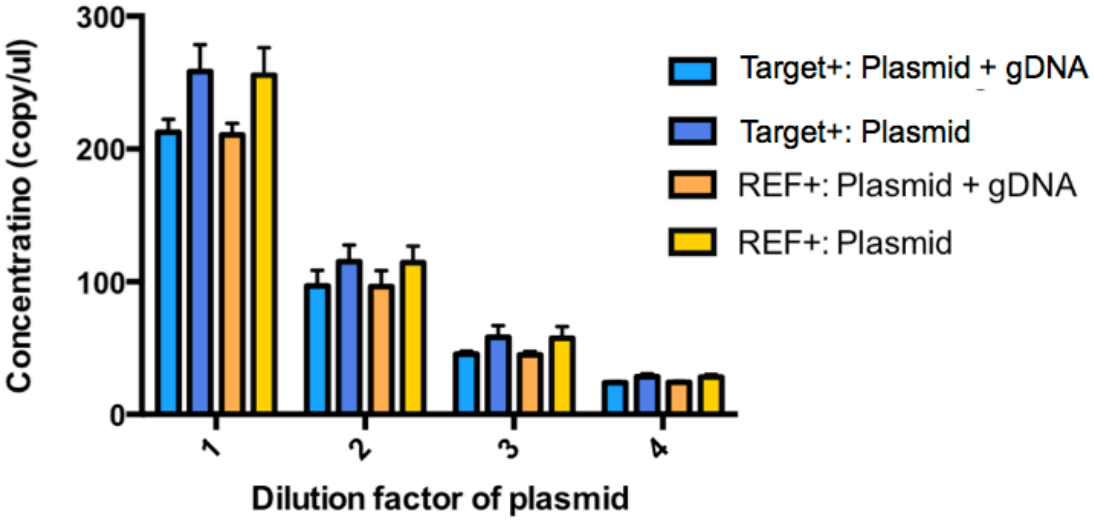

Concentratino (copy/ul)
Dilution factor of plasmid
Target+: Plasmid + gDNA
Target+: Plasmid
REF+: Plasmid + gDNA
REF+: Plasmid

Figure 10 | Result from Experiments Concerning Formation of Single Positive Droplets. A. Amplitude plot from SupT1 gDNA with no HIV sequences integrated. No nonspecific amplification was observed. B. Percentage of single positive droplets from Plasmid_WT/SupT1 mixture. Similar percentages of single positive droplets were observed in Plasmid_WT/SupT1 mixture and Plasmid_WT in all 4 dilution levels. C. Total Target positive droplets and REF positive droplets from Plasmid_WT/SupT1 mixture and Plasmid_WT in all 4 dilution levels. An inhibition level of 18% was observed in all dilution levels for both WT probe and REF probe.

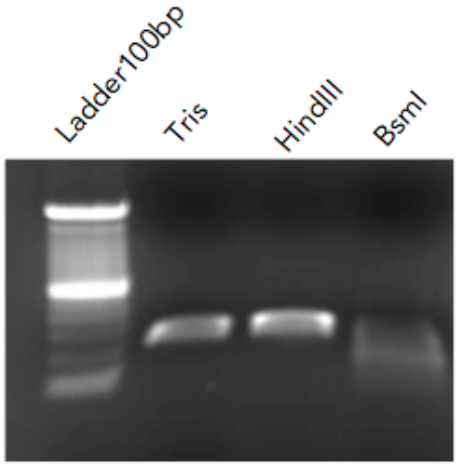


Figure 11 | Digestion comparison between HindIII and BsmI. Amplicons from PCR using Plasmid_WT as template and HIV_Ref_F and HIV_WT_R as primers were digested with Tris (control), HindIII or BsmI. There are one visible band from Tris and HindIII digested samples. But the band from BsmI digested sample disappear, indicating digestion of the amplicon.

Linearity

SupT1-DHIV3 was diluted into a five level dilution to evaluate the analytical linearity for both probes in Double-Amplicon Assay. The signal level range coves possible analysis range (50 – 400 copy/ul). 12 repeats of each dilution level were analyzed on the same plate to address aliquoting variability. 3 repeats of each dilution level were analyzed on one plate and four plates were set up in total to address between-run variability. The linear regression performed with Prism 6 is listed as Figure 10. The slopes of two lines were tested and gave a P value of 0.5116, meaning that the differences between the slopes are not significant. The pooled slope equals 389.131. The Y intercepts were tested also. A P value of 0.2456 was given, meaning that the Y intercepts of the two lines are not significantly different. The pooled intercept equals -41.9155. More statistic results are listed in Table 14. Such analysis has shown that both WT probe and REF probe have great linearity with similar linear regression equations.

	Slope	Y-intercept when X=0.0	R square
Target Probe	393.3 ± 9.277	-46.52 ± 6.154	0.9687
REF Probe	385.0 ± 8.679	-37.31 ± 5.757	0.9714
Pooled	389.131	-41.9155	

Table 14| Results from linearity analysis of Target probe and REF probe.

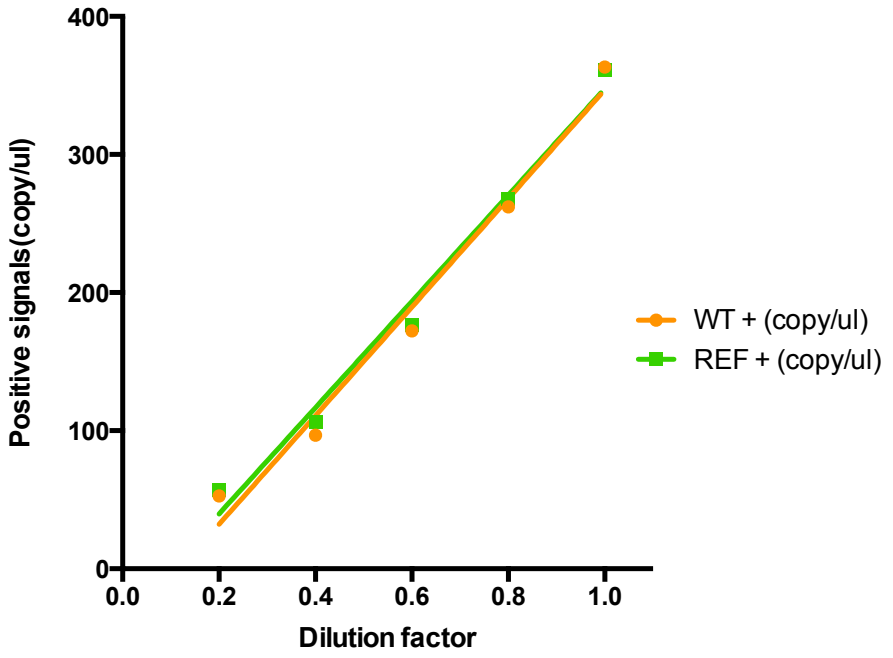


Figure 12 | Analytical linearity of WT probe and REF probe for Double-Amplicon Assay.

Reproducibility

The evaluation of reproducibility was performed with the same data set since there were four repeats of each dilution level analyzed on the same plate to evaluate within-run variability, and three plates with the same setup were analyzed to establish between-run variability.

Biostatistician, Anqi Cheng, described statistical method as the following:

$$Variance\ decomposition:$$

$$Total\ variance = Between\ run\ Variance + Within\ run\ Variance$$

$$N - 1 = (N - K) + (K - 1)$$

In this case, for each dilution level, the total sample number N=12, number of groups (run) K=4, replications m=3.

$$Var(between\ run) = \frac{\sum_{j=1}^{K} m(\bar{Y}_j - \bar{Y})^2}{K-1}, \qquad CV(between\ run) = \frac{SD(between\ run)}{\bar{Y}}$$

$$Var(within\ run) = \sum_{j=1}^{K} \frac{\sum_{i=1}^{m} (Y_{ji} - \bar{Y}_j)^2}{N-K}, \quad CV(within\ run) = average\left(\frac{SD(Y_j)}{mean(Y_j)}\right)$$

Dilution Factor	0.2	0.4	0.6	0.8	1
Within-run CVs	0.58	0.34	1.07	1.67	-1.60
Between-run CVs	0.66	0.55	1.14	0.91	-5.22

Table 15 | Within-run and between-run CVs for Double-Amplicon Assay.

As conclusion, this assay demonstrated good reproducibility in the analyze range.

Limit of Blank (LoB) Estimation

Limit of Blank is defined as "the highest apparent analyte concentration expected to be found when replicates of a blank sample containing no analyte are tested". Normally a 5% type I error risk is desired so only 5% of false positivity will exceed the LoB. LoB is a reasonable starting point for estimating the Limit of Detection.[54]

To collect 60 measurements for LoB estimation, SupT1-DHIV3 (known wild type sample) was diluted into a five level dilution. Twelve repeats of each dilution level were analyzed on the

same plate to address aliquoting variability. Four repeats of each dilution level were analyzed on one plate and three plates were set up in total to address between-run variability. A scatter plot of false mutation rate and REF + signal level detected from SupT1-DHIV3 is showed in Figure 13.

The LoB established is 9.06**%**.

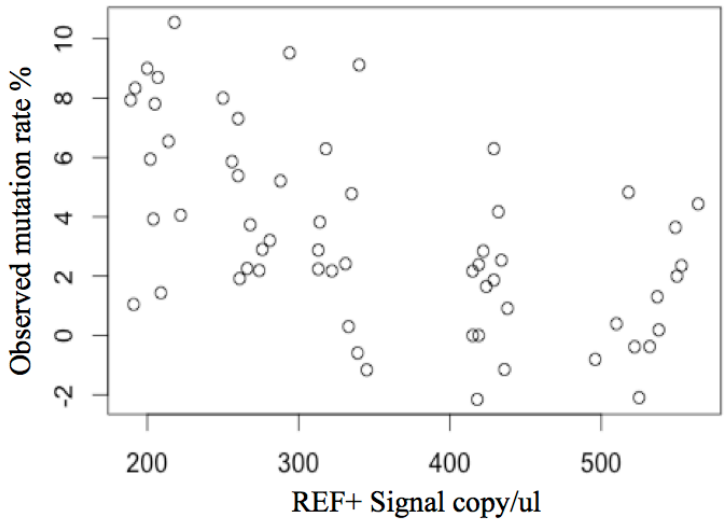


Figure 13 | Scatter plot of false mutation rate detected by Double-Amplicon Assay from SupT1-DHIV3 sample.

The samples from different plates were combined to calculate the LoB. The most common treatment, a nonparametric method was applied, since it makes no assumptions on the underlying data distribution. The desired Type I error risk is 0.05, thus the corresponding percentile is 0.95. The rank position is 57.5. By interpolating between values of the rank position 57 and 58, the resulting LoB estimate is 9.06**%**. (Anqi Cheng)

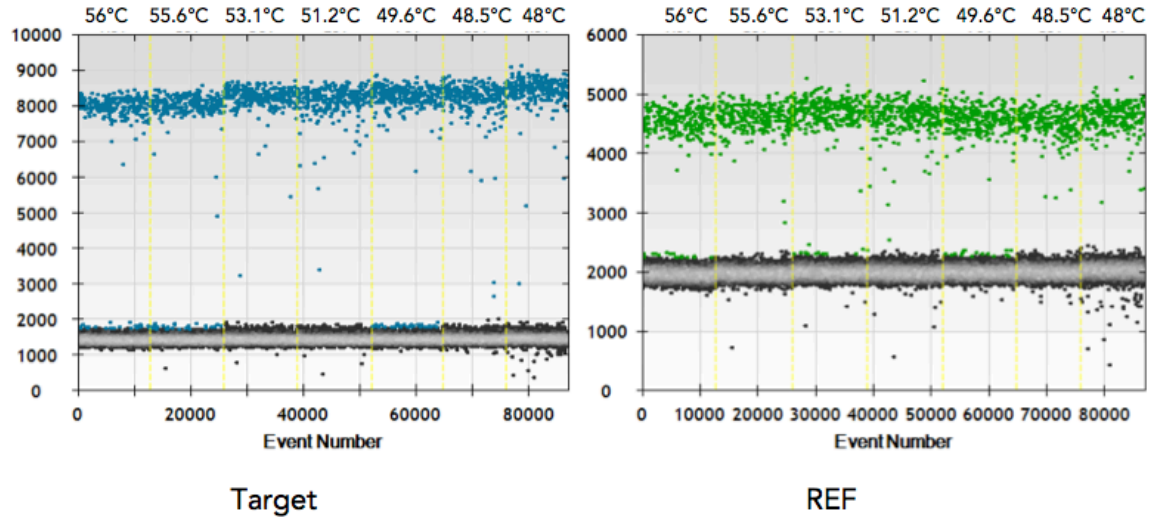
Single-Amplicon Assay B

A single-amplicon assay was developed and validated with both plasmid and cellular samples. Limit of Blank and Limit of Detection for this assay was estimated.

Assay Development on Mutant Plasmid Samples

The same plasmid samples used in validation of Double-Amplicon Assay were used to validate Single-Amplicon Assay B. One primer set and two TaqMan MGB probe were designed and tested with the wild-type plasmid (Plasmid-WT). The gradient ddPCR results for both designs are shown in Figure 14(A). Single-Amplicon Assay B showed similar amplitude level within the temperature range when using Plasmid_WT as template. Non-specific amplification was not observed. To optimize the separation between droplet group of wild type sequences and 1bp deletion sequences, another gradient PCR with a mixture of Plasmid_WT and Plasmid_Mut1 as template was performed. As showed in Figure 14(B), 57.2°C gave sufficient separation between WT droplets and 1bp deletion droplets and high amplitude for REF Probe at the same time. Therefore, for Single-Amplicon Assay B, 57.2°C was chosen for all subsequent experiments as annealing temperature. Different mutant plasmids were observed as different droplet groups, as shown in Figure 15 With such assay design all REF single positive droplets are identified as containing mutant sequences (green and orange droplets in Fig5). (New figure needed)

A.



B.

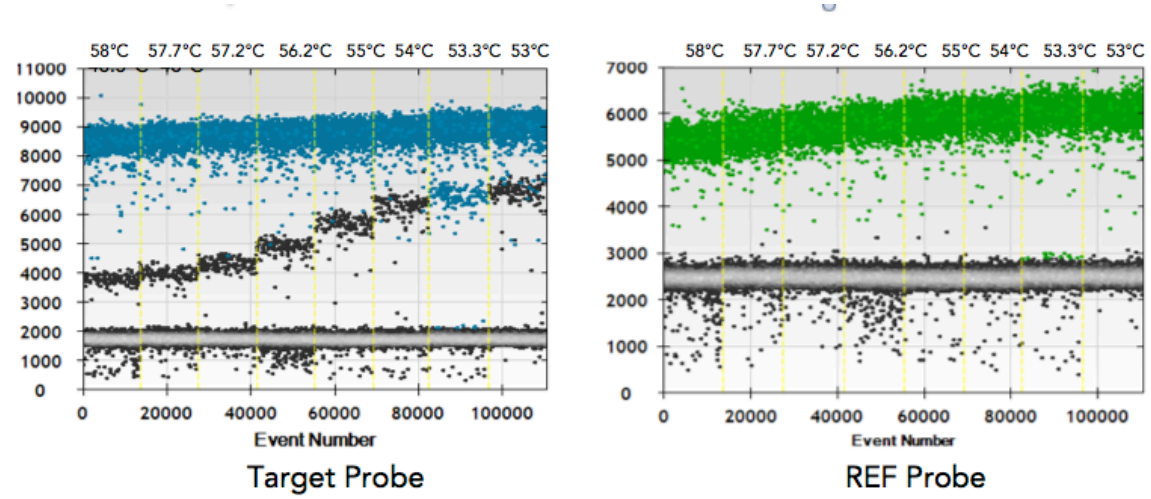


Figure 14 | Gradient PCR results in 1-D amplitude plot. (A) Template: Plasmid_WT. All temperatures show similar amplitude. (B) Template: Mixture of Plasmid_WT and Plasmid_Mut1. At 57.2°C the different plasmids were separated nicely by Target probe.

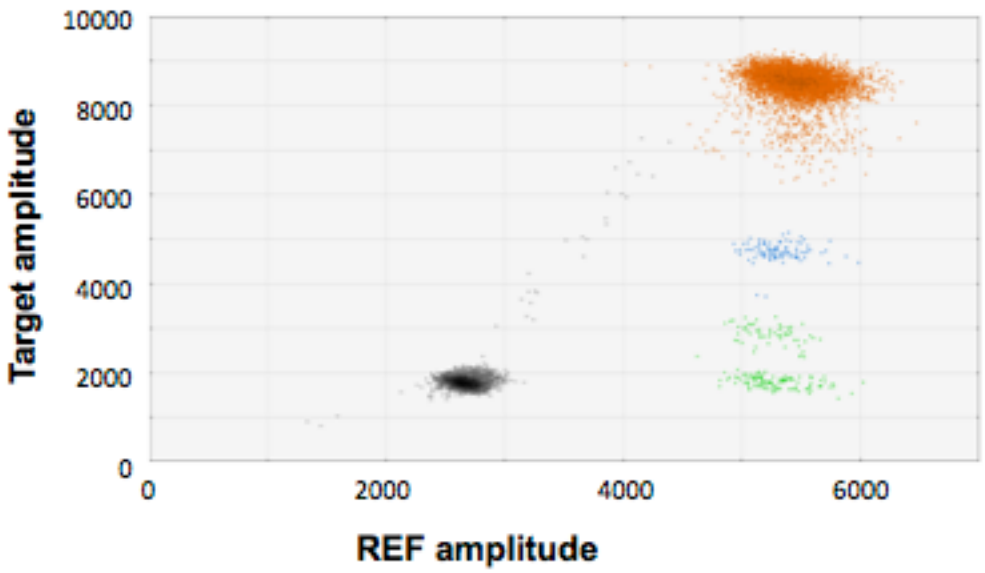


Figure 15 | Different droplet groups observed with plasmid samples in Single-Amplicon Assay B. WT: droplet group observed from Plasmid_WT (orange). Mut1: droplet group (orange) observed from Plasmid_Mut1 (1bp deletion) presents a medium amplitude due to less-efficient binding of the probe to the 1bp deletion sequence. Larger deletions (Mut2-7) abrogate probe binding and exhibit very low amplitude droplets (green).

Detecting Low-Level Mutants from A WT Background

Plasmid_Mut2 was selected as the sample here because it can represent the average of all mutation possibly observed. Five different low level Plasmid-Mut2 (2bp deletion) dilutions with constant Plasmid-WT were mixed and analyzed, the results of which were defined as duplex mutant sequences. The same Plasmid-Mut2 dilutions were analyzed simultaneously, and the result was defined as the singleplex, the readout of which should equal to the Plasmid_Mut2 level in the mixture. To determine if low-level mutants were detected by this assay, the two sets of results were compared. The estimation of repeatability (coefficient of variance) was

performed with the same data sets. The predicted rate of falsely identifying a mutant sequence, background, was estimated with data collected from Plasmid-WT only wells.

The dilution was mixed with a constant Plasmid-WT level of 490 copy/ul. From previous experiments, a low background level for this assay was observed. So as this experiment was set up, an extremely low level dilution of Plasmid_Mut2 was made to evaluate the ability of this assay to detect low level samples. In the mixture, the highest Plasmid_Mut2 level presents a mutation rate about 0.1%. The predicted rate of falsely identified mutant sequence is about 0% for such Plasmid-WT input. (Figure 16)

	Plasmid_WT	intraassay CV	Background
Mut_2	490 copy/ul	5.4%	0

Table 16 | Reproducibility and background when detecting mutants from massive WT.

For mutants with such low level, Single Amplicon Assay can still detect mutation in a semi-quantitive manner, since the results of the mixture present the same trend with the results from the singleplex wells, but has a tendency of underestimating the results. For the Plasmid-Mut2 dilution with the lowest level, in both duplex and singleplex wells, only one out of three wells had detected mutations, which means the instrument limit of detection was reached. The lowest Plasmid_Mut2 level detected (the second lowest level dilution in this experiment) mixed with Plasmid_WT gives a mutation rate level around 0.02%.

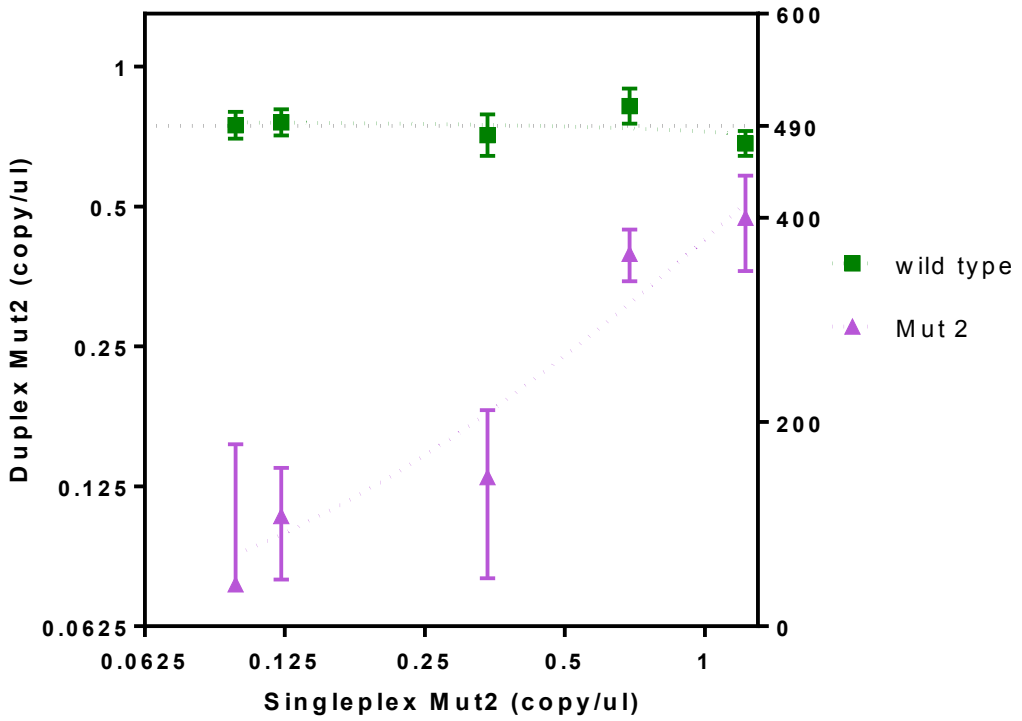


Figure 16 | Detecting extremely low-level mutants from massive WT background. Mutant sequences observed from Plasmid_WT and Plasmid-Mut2 mixture were plotted to the left Y-axis, while the results from Plasmid-Mut2 only wells (mutant sequences input) were plotted to X-axis. The Plasmid-WT level was plotted to the fight T-axis.

Linearity, Repeatability and Limit of Blank

SupT1-MegaTAL Site WT was diluted into a 5 level dilution to evaluate the analytical linearity for both probes in Double-Amplicon Assay. The signal level range coves possible analysis range (100 – 500 copy/ul). Most cellular samples will for into this range after digestion and dilution. Also, this range assures that enough molecules are analyzed by the assay. 12 repeats of each dilution level were analyzed on the same plate to address aliquoting variability. 3 repeats of each dilution level were analyzed on one plate and four plates were set up in total to address between-run variability.

Linearity

The linear regression performed with Prism 6 is listed as Figure 17. The slopes of two lines were tested and gave a P value of 0.5116, meaning that the differences between the slopes are not significant. The average slope was 471.1. The Y intercepts were tested also, and the Y intercepts of the two lines are not significantly different. The pooled intercept equals 1.479, which is very close to 0. More statistics are listed in Table 17. This analysis has shown that both Target probe and REF probe have great linearity with similar linear regression equations.

	Slope	Y-intercept when X=0.0	R square
Target Probe	470.4 ± 6.729	1.558 ± 4.463	0.9883
REF Probe	471.7 ± 6.634	1.400 ± 4.400	0.9887
Average	471.1	1.479	

Table 17 | Within-run and between-run CVs for Single-Amplicon Assay B

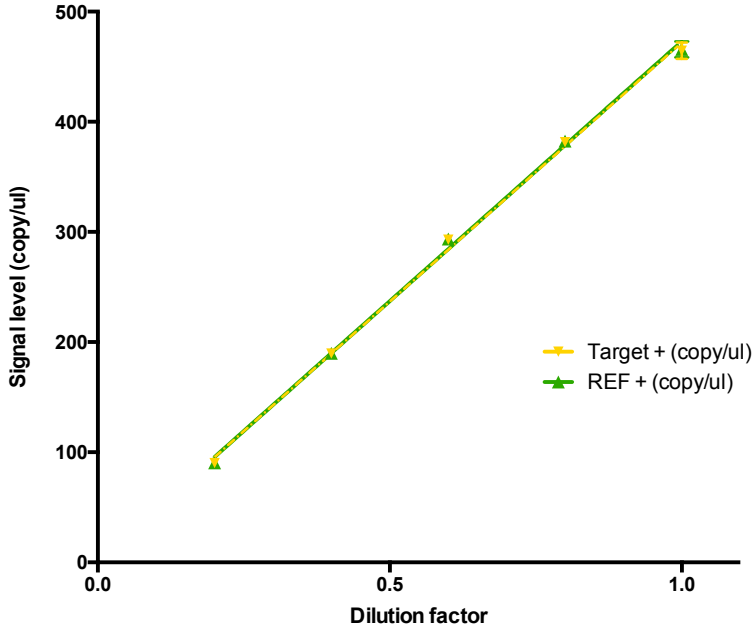


Figure 17 | Analytical linearity of WT probe and REF probe for Single-Amplicon Assay B.

A.

Dilution Factor	0.2	0.4	0.6	0.8	1
Target	5.82%	6.33%	1.90%	2.11%	5.00%
REF	5.56%	6.30%	1.94%	2.01%	4.90%

B.

Dilution Factor	0.2	0.4	0.6	0.8	1
Target	1.15%	2.00%	3.44%	3.48%	6.46%
REF	1.39%	1.79%	3.51%	3.40%	6.23%

Table 18 | Repeatability (CVs) for Single-Amplicon Assay B. (A) Within-run CVs; (B) Between-run CVs.

Repeatability

The evaluation of reproducibility was performed with the same data set since there were 4 repeats of each dilution level analyzed on the same plate to evaluate within-run variability, and 3 plates of the same set-up were analyzed to establish between-run variability. The analysis was performed the same way as previously with Double Amplicon Assay

As conclusion, this assay demonstrated good reproducibility in the analyze range.

Limit of Blank

The analysis was performed with the same statistical method as in Double-Amplicon Assay. The rank position is 57.5. By interpolating between values of the rank position 57 and 58, the resulting LoB estimate is 0.56%.

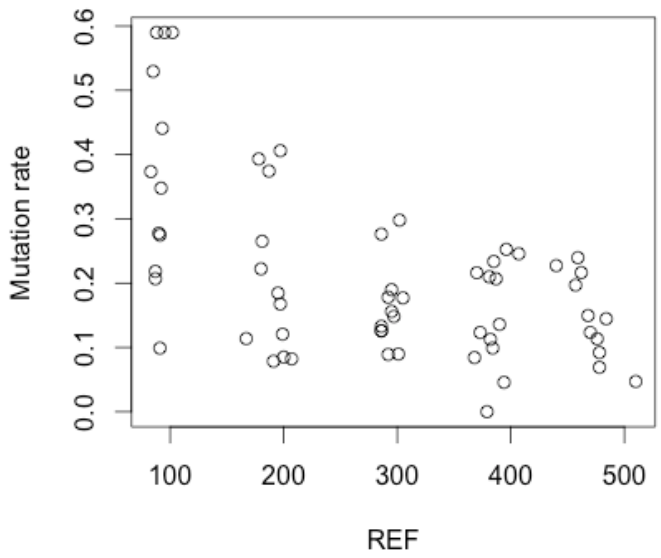


Figure 18 | Scatter plot of false mutation rate detected from SupT1-HIV-WT sample.

The samples from different plates were combined to calculate the LoB. The most common treatment, a nonparametric method was applied, since it makes no assumptions on the

underlying data distribution. The desired Type I error risk is 0.05, thus the corresponding percentile is 0.95. The rank position is 57.5. By interpolating between values of the rank position 57 and 58, the resulting LoB estimate is 0.56%. (Anqi Cheng)

Limit of Detection

Limit of Detection is defined as "the lowest analyte concentration likely to be reliably distinguished from the LoB and at which detection is feasible." Again, since the type I error risk desired for LoB estimation is usually 5%, this means only 5% of low concentration samples will not produce a result below the LoB and identified as false negative.[54]

The LoD estimated for this assay is 1.06%. (Anqi Cheng)

Limit of Quantitation

Limit of Quantitation is "the lowest concentration at which the analyte can not only be reliably detected but at which some predefined goals for bias and imprecision are met." The imprecision goal set for this assay was 20%, which is the commonly defined "functional sensitivity", also a measure of the precision of an assay at it's low analytical range.[54]

The LoQ estimated for this assay is 2.19%. (Anqi Cheng)

Mutation Detection in MegaTAL Treated Samples

All megaTAL treated samples and control samples were analyzed by both assays. All samples were diluted before analysis so total DNA input would fall into the recommended range. Samples were analyzed in triplicate, and two separated runs were performed. The results and CVs are shown below. (Fig.17, Table.19).

(A)

	Mutation Rate	Within-run CV	Between-run CV
DHIV3+ 6.5	43.15%	3.78%	1.80%
pDHIV3 + 6.5	62.45%	2.12%	2.00%
pDHIV3 + 6.5 +Trex2	55.65%	3.52%	3.72%
pDHIV3 + 7.5	61.24%	1.82%	0.96%
pDHIV3 + 7.5 + Trex2	53.29%	2.68%	1.80%

(B)

	Mutation Rate	Within-run CV	Between-run CV
DHIV3	1.62%	16.35%	6.10%
DHIV3+ 6.5	3.98%	16.46%	31.3%
pDHIV3	0	NA	NA
pDHIV3 + 6.5	14.22%	5.23%	6.49%
pDHIV3 + 6.5 +Trex2	33.05%	3.21%	2.34%
pDHIV3 + 7.5	14.07%	3.48%	8.81%
pDHIV3 + 7.5 + Trex2	27.20%	5.88%	1.60%

Table 19 | Mutation rate estimated by ddPCR, within-run and between-run CVs for all megaTAL treated samples. (A) Double-Amplicon Assay; (B) Single-Amplicon Assay A.

Method Comparison With Surveyor Assay, Clonal sequencing and Illumina Sequencing

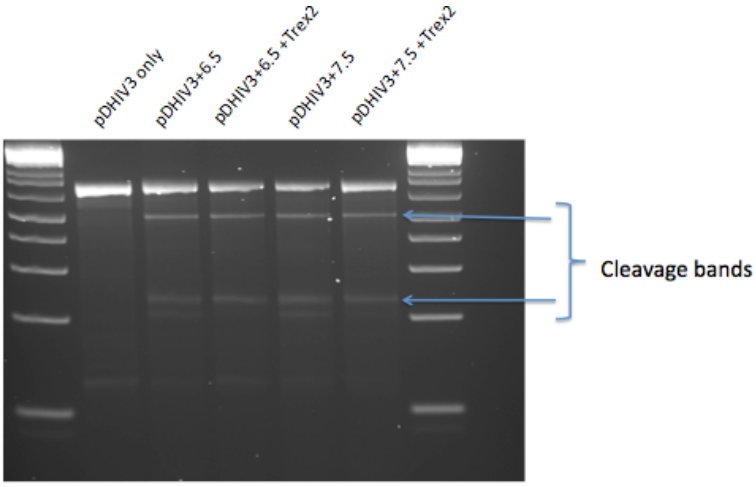
Surveyor Assay

The 293T control sample and all treated cell line samples were analyzed for mutation rates through Surveyor assay. For SupT1-DHIV3 + 6.5, no visible cleavage bands were observed. For control sample and megaTAL treated samples obtained form 293T cells, two different quantities of DNA inputs (400ng and 200ng) were analyzed. An image from the gel with 200ng DNA input is showed in Figure 11A. An example of gel analysis is shown in Figure 20B. The results from the two experiments are listed in Table 20.

	400ng	200ng	Mean
pDHIV3 + 6.5	10.24%	10.12%	10.18%
pDHIV3 + 6.5 + Trex2	13.42%	12.16%	12.79%
pDHIV3 + 7.5	19.34%	14.21%	16.78%
pDHIV3 + 7.5 + Trex2	9.09%	11.43%	10.26

Table 20 | Mutation rate estimated with Surveyor assay.

A.



B.

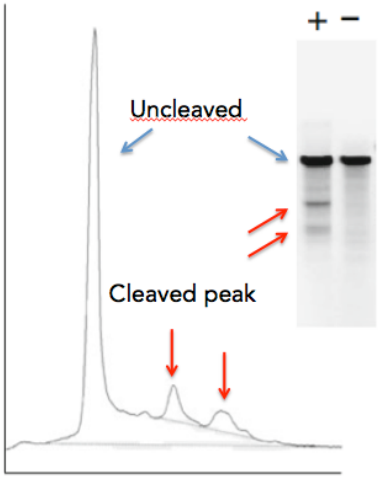


Figure 20 | Surveyor assay gel image and analysis. A. Gel image from Surveyor assay with 200ng DNA input. There are two visible cleavage bands and one uncleaved band from megaTAL treated samples, but only one uncleaved band from control sample. B. An example of gel analysis with ImageJ. The fraction cleaved is calculated as the sum of two cleaved peaks divided by total density. The percentage of mutation can be calculated with the following equation:

$$\% \text{ Mutation} = 100 \times (1 - (1 - \text{fraction cleaved})^{1/2})$$

Clonal sequencing

All megaTAL treated samples were amplified, and single amplicons were separated and sequenced individually by Sanger sequencing. The results are listed in Table 21.

	Number of Amplicons	Mutation Detected	Mutation Rate
DHIV3+ 6.5	50	4	8%
pDHIV3 + 6.5	31	5	16.13%
pDHIV3 + 6.5 + Trex2	27	12	44.44%
pDHIV3 + 7.5	19	3	15.79%
pDHIV3 + 7.5 + Trex2	40	10	25%

Table 21 | Results form clonal sequencing. For each sample, the number of amplicons sequenced successfully is listed in the first column, the number of mutations detected is listed in the second column, and the mutation rate established is listed in the third column.

Illumina Sequencing

All sequencing data was produced with a 2*300 paired end read kit on the MiSeq. Sequencing output was processed in Geneious to filter out all portions of reads that contain any base of Q30 or less, any ambiguous calls and any reads of length <200bp after filtering. The read numbers previous to and after this trimming step are listed in Table 22.

All mutations affecting a 20bp range centered at “central four” were considered as results from megaTAL treatment. Results from analysis of this target site sequencing are listed in Table 23. The deletion detected in SupT1-DHIV3 is a 32 bp deletion covering they megaTAL recognition site. In 293T – pDHIV3 samples, megaTAL 6.5 showed higher modification efficiency then megaTAL 7.5, and Trex2 increased the mutation rate. In some sequences, besides deletions as nuclease loss, modifications also appeared around “Central four” simultaneously. Such so-existence of deletion and modification is more common in Trex2 treated samples. A 1bp modification was observed in all samples with similar rate, indicating a naturally occurred single-nucleotide polymorphism.

Occasionally, when a mutation affects other binding regions beside Target probe binding region, ddPCR will not be about to detect it, therefore under estimating the mutation rate. The percentages of such mutations in each sample are listed in Table 24.

	Raw Number of reads	After trimming and Filtering	Mapped to Reference Sequence
SupT1-DHIV3 (Negative Control)	602k	2.2k	80
SupT1-DHIV3 + 6.5	306k	1.45k	1186
293T + pDHIV3 (Negative Control)	151k	17k	17123
pDHIV3 + 6.5	435k	25k	24654
pDHIV3 + 6.5 + Trex2	550k	32k	31511
pDHIV3 + 7.5	203k	8.8k	8711
pDHIV3 + 7.5 + Trex2	151k	6.2k	6098

Table 22 | Number of reads for each sample sequenced by Illumina. Number of reads were trimmed.

	SupT1-DHIV3		293T - pDHIV3				
	SupT1-DHIV3	DHIV3 + 6.5	pDHIV3	pDHIV3 + 6.5	pDHIV3 + 6.5 + Trex2	pDHIV3 + 7.5	pDHIV3 + 7.5 + Trex2
Total reads	80	1186	17123	24654	31511	8711	6098
Mutation rate detected	**3.8%**	**5.6%**	**2.4%**	**24.2%**	**47.1%**	**20.0%**	**26.5%**

Table 23 | Illumina sequencing results. Target site Illumina sequencing analysis results. “Central four” are underlined.

	SupT1-DHIV3		293T - pDHIV3				
	SupT1-DHIV3	DHIV3 + 6.5	pDHIV3	pDHIV3 + 6.5	pDHIV3 + 6.5 + Trex2	pDHIV3 + 7.5	pDHIV3 + 7.5 + Trex2
Mutation rate	**3.8%**	**5.6%**	**2.7%**	**24.2%**	**47.1%**	**20.0%**	**26.5%**
Mutations affecting REF probe	0	0.2%	0	1.6%	1.8%	0.7%	0.7%
Mutations affecting Fwd Primer	0	0	0	0.5%	0.9%	0.2%	0.5%
Mutations affecting Rev Primer	2.5%	0.2%	0	2.2%	2.4%	1.9%	1.4%
Mutations rate differences	2.5%	0.3%	0	3.7%	4.4%	2.5%	2.3%

Table 24 | Mutations affecting REF probe/Fwd Primer/Rev primer binding regions: mutations that affecting other primer/probe binding regions in Single-Amplicon Assay B. Mutation rate difference: Mutations that theoretically cannot be detected by Single-Amplicon Assay B due to mutations in REF probe or two primers binding regions. Mutation rate differences – SNPs: Mutations that theoretically cannot be detected by Single-Amplicon Assay B due to SNPs in REF probe or two primers binding regions

Result Comparison

Results from other methods and two ddPCR assays are listed in Table 25.

	Surveyor	Clonal sequencing	Illumina	Double-Amplicon Assay	Single-Amplicon Assay B
DHIV3	Cleavage Not Visible	Not Performed	2.4%	0	1.62%
DHIV3 + 6.5	Band Not Visible	8% (4/50)	3.7%	43.15%	3.98%
pDHIV3	Cleavage Not Visible	Not Performed	0.2%	0	0
pDHIV3 + 6.5	10.18%	16.13% (5/31)	22.5%	62.45%	14.22%
pDHIV3 + 6.5 + Trex2	12.79%	44.44% (12/27)	45.9%	55.65%	33.05%
pDHIV3 + 7.5	16.78%	15.79% (3/19)	17.9%	61.24%	14.07%
pDHIV3 + 7.5 + Trex2	10.26%	25% (10/40)	24.9%	53.29%	27.20%

Table 25 | Result comparison from different methods.

Chapter 6 Discussion

In my thesis, three digital droplet polymerase chain amplification (ddPCR) assays -- one double-amplicon and two single-amplicon assays -- were developed to detect megaTAL-induced HIV *pol* mutations. Of the three assays, only the double-amplicon assay and one of the two single-amplicon assays demonstrated promising results using plasmids and were further validated using cell-line derived (cellular) samples. However, the double-amplicon assay was not reliable with cellular samples while the single amplicon assay (Single-Amplicon Assay B) showed greater sensitivity and repeatability and detected mutants with differences as small as 1bp. For a mixed mutant and wild-type population, the limit of blank (LoB) for this assay for detecting the mutant was 0.56% and the limit of detection (LoD) was 1.06%, which was significantly lower than either of the two comparators -- bulk sanger sequencing (20%)[55] or Surveyor assay (10%)[56].

Analyzing cellular samples

As shown in Figure 9, when analyzing cellular samples with wild type HIV sequences, the Double-Amplicon Assay gave some unexpected single-probe positive droplets, which affected the interpretation.

I hypothesized three possible mechanisms behind the formation of single-positive droplets. The first one is non-specific amplifications within the human genome, which was ruled out by analyzing SupT1 gDNA with Double-Amplicon Assay (Fig 10(A)). The second possible mechanism is unbalanced interference from gDNA for each probe. I have statistically shown that the levels of inhibition (18%, P = 0.34) from gDNA to two probes was similar and gDNA interference was not the reason for single-positive droplets (single Target positive: P = 0.25~~04~~,

single REF positive: P = 0.14). The last possible mechanism is DNA fragmentation. As shown in Figure 1(B), there was a 94bp gap between the two amplicons. If DNA was fragmented during extraction, there is a possibility that the two regions will be unlinked, forming droplets that contain only one of the two regions during digitalization, which will become single-positive droplets when analyzed. In the original design, the single REF positive droplets were considered as mutants. Therefore, I concluded that fragmentation increased the false mutation rate of this assay significantly.

To minimize the influence of random DNA fragmentation, a restriction enzyme, BsmI, was used in sample preparation to unlink the two regions purposefully. Since now the two amplicons are randomized into droplets separately, the missing Target signal is considered as mutant. The mutation rate was calculated using following equation:

$$Mutation\ Rate\ (\%) = \frac{REF\ (copy/ul) - Target(copy/ul)}{REF(copy/ul)} * 100\%$$

Comparison Of The Two ddPCR Assays

The two designs share the same mutation reporting principle. When a mutation occurs around an enzyme-cleavage site, the target probe will not bind to the DNA fragments; therefore there will be an absence of Target probe-positive signal.

However, Single-Amplicon Assay B has shown to be superior analytically, especially when it comes to LoB: 9.06% for Double-Amplicon Assay and 0.56% for Single-Amplicon Assay B. The reason for such a large LoB difference may be the primer efficiency. For the Double-Amplicon assay, the primers for the amplicon with REF probe binding region may have slightly higher binding efficiency. As a consequence, even if there were equal molecule numbers of two regions in the assay; the REF probe might still produce more signals. Since the absence of

Target probe signal compared to REF probe signal was defined as mutant, this assay will over estimate the mutation rate.

The threshold setting in Single-Amplicon Assay B is more intuitive than for Double-Amplicon assay. In Single-Amplicon Assay B, mutant droplet clusters were visible on the 2-D plot QuantaSoft provides, and the mutant concentration were achieved directly from the software by gating the droplet clusters; while in the Double-Amplicon Assay, there were several clusters in the amplitude plots that imply mutants, which may cause confusion. Also, since the separation between wild type droplets and 1bp deletion droplets was not optimized for the Double-Amplicon Assay, the identification of mutation with small differences was not as reliable as for the Single-Amplicon Assay B, which was optimized for separation between the two groups.

In conclusion, Single-Amplicon Assay B was the demonstrably better ddPCR assay for detecting megaTAL induced *pol* gene mutations.

Methods Comparison

Three comparator mutation detection methods were used to analyze the same research samples: Surveyor assay, clonal sequencing and Illumina sequencing. These methods are most commonly used in research labs for detecting unknown mutations in certain gene regions. Since next-generation sequencing is widely accepted as the gold standard in this field, the results from all methods used in this project, including the two ddPCR assays developed, were compared to the results from Illumina sequencing.

Surveyor Assay

Surveyor assay is a widely used mutation detection method in research labs. It is rapid and can detect mutations as small as single-nucleotide polymorphisms without batch-to-batch

optimization. We used it to analyze all research samples. As shown in Table 24, the results from Surveyor assay were lower than the results from Ilumina sequencing over all.

To understand why Surveyor assay has a tendency of underestimating mutation rate, we need to look into the principle of it first (Figure 21).

In the first step of the reaction, one mutant DNA strand need to form a heteroduplex with a wild type DNA strand so the enzyme would recognize the mismatch in a later step. There are two possible reasons that can lead to an underestimated mutation rate. If the mutation rate for the sample analyzed is very high, there is a higher chance that mutant strands will form homoduplexes with each other such that the enzyme will not recognize and cleave the target sequence. Secondly, if the mutations harbored in the sample are large deletions or insertions, as is the case in our research sample pDHIV3 + 6.5 + Trex2, then the mutant strand will not be able to form a heterodueplex with a wild-type strand because of the large differences between them; therefore the surveyor enzyme will not be able to cleave the target sequence.

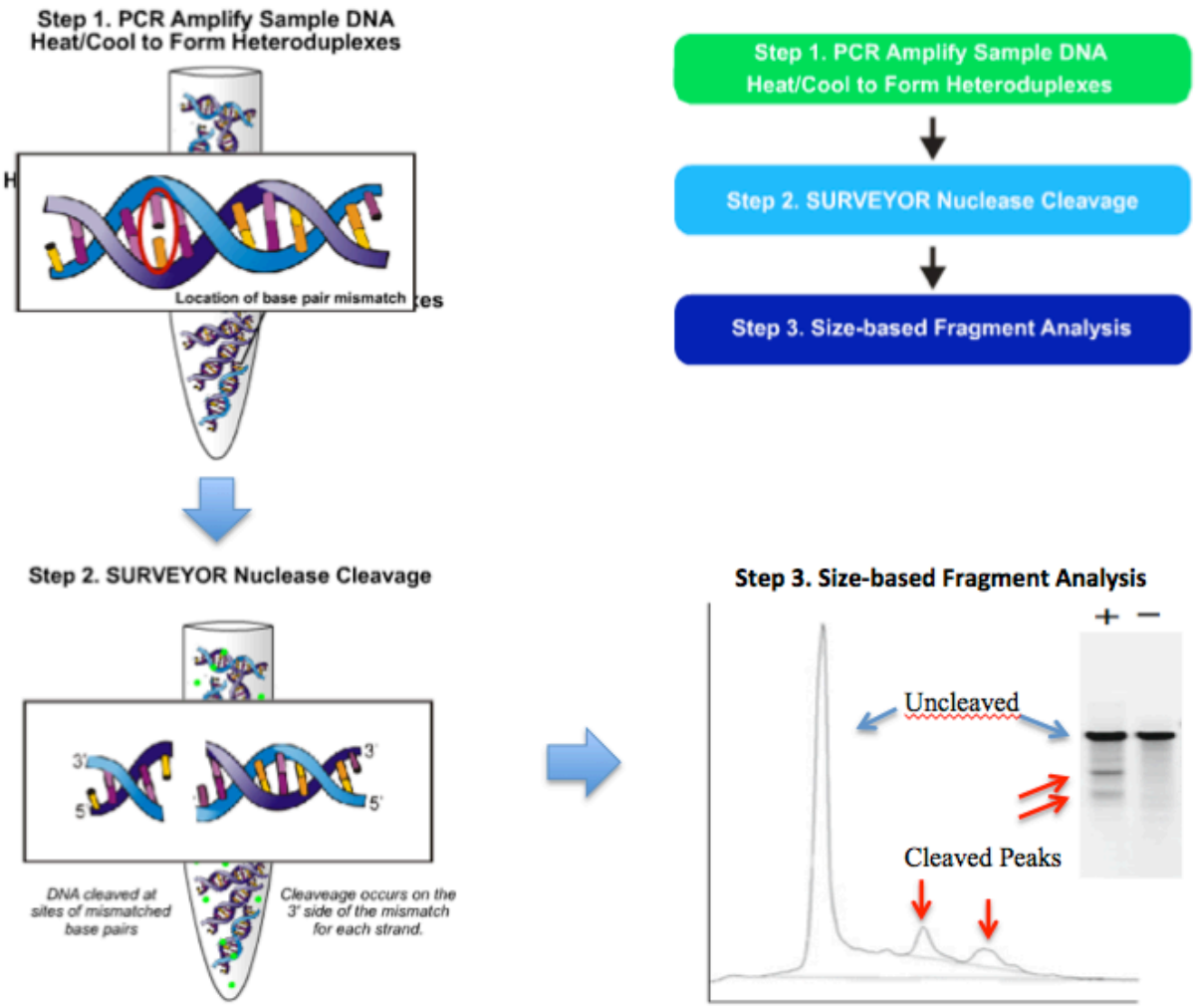


Figure 21 | Surveyor assay steps. Step 1: Mutant and wild type strands are hybridized by heating and cooling the mixture to form hetero- and homoduplexes. Step 2: The hetero- and homoduplex mixture is treated with Surveyor nuclease, which recognizes mismatches in heteroduplexes and create a double-strand break. Step 3: The DNA fragments are analyzed by size-based fragment analysis, such as agarose gel electrophoresis, the image of which is analyzed with software such as ImageJ.

When analyzing the gel with software, bias can come from two sources. The uncleaved bands on the gel might have extremely high density if the loaded DNA was not in a proper input range,

which may overwhelm the peaks of other bands, leading to a lower fraction cleaved. Also, since the threshold setting is normally performed manually, the result of the density of each peak is arbitrary. These are also the two main reasons that gel-based methods are usually considered semi-quantitative.

In DHIV3 + 6.5, since the integrated rate for this cell line is so low, no bands were visible on the gel for analysis. Even if the DNA loaded was enough to show bands, I suspect that we cannot see the two cleaved bands, since the mutation rate estimated by Illumina for this sample is 2.9%.

In a study comparing Surveyor assay with another similar mutation detection method, the Surveyor assay was able to detect from 10% to 90% of mutant DNA, but could not detect 5% mutant DNA, which suggests that Surveyor has a limit of detection between 5% and 10% mutation rate. [57]

In conclusion, although the Surveyor assay is a simple and fast assay for detecting from 10% to 90% of mutant DNA semi-quantitatively, the ddPCR offered better sensitivity and absolute quantification.

Clonal sequencing

Clonal sequencing, also known as amplicon analysis, is a Sanger sequencing based method for mutation detections and is capable of detecting all types of mutations, making it suitable for highly variant samples. All research samples were analyzed with this method, and the results are listed in Table 20 and 24. As showed in Table 24, the results from clonal sequencing were very close to the results from Illumina sequencing.

However, clonal sequencing has its own drawbacks. First is the limited sample numbers that can be analyzed. As shown, the largest sample size for the samples is 50. Statistically, a larger

sample size generates results that are more convicting and reliable. Comparing to our ddPCR assay, Single-Amplicon Assay B, which can analyze as many as 10,000 molecules in one well, the sample size of clonal sequencing is really very small.

Also, clonal sequencing requires multiple cloning and purification steps before Sanger sequencing steps, which leads to longer turn around time.

As a conclusion, although clonal sequencing gives specific non-quantitative information about mutation types, the reliability of this method is compromised by the small sample size. By comparison, ddPCR offers reliable absolute quantification results with short turnaround time.

Illumina Sequencing

Next-generation sequencing (NGS) is a powerful tool that can be utilized in a wide collection of genetic research topics. In mutation detection of a targeted region, NGS offers highly sensitive detection with very low background while providing specific information about the mutations detected. NGS is now considered a gold standard in the diagnostic field. However, Illumina sequencing is not suitable for routine performance in research labs for a few reasons. First, even though the cost for NGS is decreasing rapidly, it is still expensive to perform routinely. Second, it requires multiple library preparation steps, which leads to a long turnaround time. Finally, the information provided by Illumina sequencing can be overwhelming during analysis if specific information is not needed. All our research samples were analyzed by Illumina sequencing. The results are listed in Table 21, 22, 23 and 24.

As listed in Table 24, when analyzing same research samples, the results of Double-Amplicon Assay were significantly higher than the results of Illumina sequencing results . The possible explanations were stated in previous section “Comparison Of The Two ddPCR Assays”.

The results from Single-Amplicon Assay B and Illumina sequencing were quite similar. However, there was a tendency for Single-Amplicon Assay B to underestimate the mutation rate.

The reason for this may lie in the positioning of primers and probes. Originally, we expected the mutations induced by megaTALs to occur closely around the cutting site (central four), within the binding region of Target probe. However, when analyzing data from Illumina sequencing, mutations that exceeded the Target probe-binding region were observed, including large deletions and insertions, in all megaTALs treated samples. Since the primers and probes have binding sites that were located very close to each other, such mutations may affect the binding of REF probe and two primers, which would lead to the absence of REF signals, regardless of the presence of HIV sequences. As a consequence, the mutation rate is underestimated in our ddPCR assay. The percentage of mutations that are theoretically under estimated is listed in Table 23 as “Mutation rate difference”.

However, if adding the “Mutation rate difference” to the results from Single-Amplicon Assay B, the results are larger than the results from Illumina sequencing. One explanation is that sometimes the mutations that occur in REF probe or primers binding regions are SNPs, which may or may not actually affect the binding of oligonucleotides. The percentage of such SNPs is listed in Table 23 as “Mutation rate difference - SNPs”.

After comparing the results from our ddPCR assay and Illumina sequencing, we concluded that Single-Amplicon Assay B offers similar results to Illumina sequencing, but has a tendency to underestimate mutation rates. However, the high performance of Illumina also cost 500 – 1,700 dollars per sample if ordered as service from biotech companies and long turnover time (one or two days for library preparation and sequencing, and more time for data analysis). When taking

all these into consideration, the significantly smaller workload for sample preparation and data analysis, the drawback of Single-Amplicon Assay B underestimating mutation rates was acceptable.

Other Mutation Detection Methods

There is a long list of mutation detection methods that are currently used in laboratories to detect mutations. Here are a few that hit the top of this list.

Several different approaches of qPCR are utilized in mutation detection. Allele-specific PCR (AS-PCR) is commonly used method in mutation detection. In allele-specific PCR, at least two oligonucleotide primers with different 3' end will be designed with the 3' ends located at the mutation spot. When the mutation occurs, the mismatch will not allow replication, while the matched primer will function normally. The appearance of a product then indicates the genotype. In one Japanese study, AS-PCR and bulk sequencing was used to analyze seven reverse transcriptase inhibitor resistance mutations and 3 protease inhibitor resistance mutations. AS-PCR not only identified all 8 mutant samples that bulk sequencing has detected, but also identified 15 additional mutants, indicating higher sensitivity. Additionally, primers for the detection of M46I/L were developed and evaluated. M46I could be detected as low as 0.04% and M46L could be detected as low as 0.03% in plasmid samples. When analyzing clinical samples, the M64I assay could detect 0.54% mutant virus, while the M46L assay could detect 4.01% mutant virus. But this assay was stated as not necessarily applicable to mutant virus quantitation.[58]

High Resolution Melt (HRM) is another PCR-based method that can be used to detect mutations. After real-time PCR amplification, a melt curve is performed in presence of

intercalating dye that gives decreasing signal while more dsDNAs denature. HIR monitors this process in real-time, and the data (melting curve) will be collected and analyzed by software. Since mutant sequences have altered melting temperature, their melting curves are distinguishable from those of wild-type sequences. HRM can detect small differences as SNPs. Several studies have utilized HRM to analyze HIV diversity due to its ability to distinguish different sequences in one single assay.[59–61]

Oligonucleotide Ligation Assay (OLA) is also commonly used in detection of known mutations. In OLA, three probes are involved: a WT probe, a Mutant probe, the 3' bases of which are situated directly over the target SNP, and a common probe that bind to the template upstream. The allelic probes are modified at their 5' ends with hapten groups for enzyme-linked immunosorbent assay (ELISA) or oligomers for size differentiation. After PCR amplification, the Mutant probe or the WT probe is ligated to a target sequence along with the common probe. Sized based methods or ELISA can detect the ligated products afterwards. OLA was used to detect protease inhibitor resistance in HIV-1. The assay has shown a LoD of 5% mutant plasmid from wild-type plasmid.[62] In another assay, OLA has demonstrated higher sensitivity then sequences.[63]

Limitation of ddPCR Assay

Like every assay, our Single-Amplicon Assay B has its own limitations, which will also apply for other ddPCR-based assays.

As discussed previously in Illumina section, Single-Amplicon Assay B had a tendency to underestimate mutation rates, especially when a sample contains high levels of mutations that interfere with the REF probe or primer binding regions. There is no way to eliminate this bias completely, but alternative designs may minimize this problem. However, given that the

amplicon length allowed for ddPCR droplet formation is quite small, there may not be a feasible alternative. Also, when designing an assay, the information about what type of mutations samples may contain is usually unknown, making it hard to decide the location of primers and probes based on such knowledge.

Like every PCR based assay, Single-Amplicon Assay B cannot provide specific information on what type of mutation is harbored in the analyzed samples. While clonal sequencing and Illumina sequencing can generate data about mutation types, Single-Amplicon Assay B can only approximately identify 1bp difference mutations from other type of mutations.

Conclusion

In this project, we sought to develop a single ddPCR-based assay platform to detect megaTAL-induced mutations in the HIV provirus *pol* gene. Three different assays were designed and evaluated, among which only one assay has shown to have suitable analytical linearity and repeatability using both plasmid and cell-line samples. The Single-Amplicon Assay B was capable of detecting mutations from as low as 1.06% (LoD), with a very low background level of 0.56% (LoB). The lowest level of mutation rate this assay can quantify is 2.19% (LoQ). Our ddPCR assay generated comparable results to Illumina sequencing, with a lower estimated cost and shorter turnaround time. The results from our ddPCR assay were easier to interpret than those from Illumina sequencing, had a better accuracy than Surveyor, and required less technical work and analysis than clonal sequencing.

References

1. Maartens G, Celum C, Lewin SR. HIV infection: epidemiology, pathogenesis, treatment, and prevention. *Lancet*. 2014;384(9939):258-271. doi:10.1016/S0140-6736(14)60164-1.

2. Gayle HD, Hill GL. Global impact of human immunodeficiency virus and AIDS. *Clin Microbiol Rev*. 2001;14(2):327-335. doi:10.1128/CMR.14.2.327-335.2001.

3. Chan DC, Kim PS. HIV Entry and Its Inhibition. *Cell*. 1998;93(5):681-684. doi:10.1016/S0092-8674(00)81430-0.

4. Zheng Y-H, Lovsin N, Peterlin BM. Newly identified host factors modulate HIV replication. *Immunol Lett*. 2005;97(2):225-234. doi:10.1016/j.imlet.2004.11.026.

5. Hallenberger S, Bosch V, Angliker H, Shaw E, Klenk HD, Garten W. Inhibition of furin-mediated cleavage activation of HIV-1 glycoprotein gp160. *Nature*. 1992;360(6402):358-361. doi:10.1038/360358a0.

6. Cooper A, García M, Petrovas C, Yamamoto T, Koup RA, Nabel GJ. HIV-1 causes CD4 cell death through DNA-dependent protein kinase during viral integration. *Nature*. 2013;498(7454):376-379. doi:10.1038/nature12274.

7. Prendergast A, Prado JG, Kang Y-H, et al. HIV-1 infection is characterized by profound depletion of CD161+ Th17 cells and gradual decline in regulatory T cells. *AIDS*. 2010;24(4):491-502. doi:10.1097/QAD.0b013e3283344895.

8. Cosgrove C, Ussher JE, Rauch A, et al. Early and nonreversible decrease of CD161++ /MAIT cells in HIV infection. *Blood*. 2013;121(6):951-961. doi:10.1182/blood-2012-06-436436.

9. Konttinen YT, Hänninen A, Fuellen G. Plasmacytoid dendritic cells, Janus-faced sentinels: progesterone, guilty or innocent? 2009. http://www.futuremedicine.com/doi/abs/10.2217/imt.09.71?url_ver=Z39.88-2003&rfr_id=ori:rid:crossref.org&rfr_dat=cr_pub%3dwww.ncbi.nlm.nih.gov&. Accessed August 25, 2014.

10. Brenchley JM, Price DA, Schacker TW, et al. Microbial translocation is a cause of systemic immune activation in chronic HIV infection. *Nat Med*. 2006;12(12):1365-1371. doi:10.1038/nm1511.

11. Sáez-Cirión A, Bacchus C, Hocqueloux L, et al. Post-treatment HIV-1 controllers with a long-term virological remission after the interruption of early initiated antiretroviral therapy ANRS VISCONTI Study. *PLoS Pathog*. 2013;9(3):e1003211. doi:10.1371/journal.ppat.1003211.

12. Porter K, Lodi S, Meyer L. Immunovirologic control 24 months after interruption of antiretroviral therapy initiated close to HIV seroconversion--reply. *JAMA Intern Med*. 2013;173(6):476-477. doi:10.1001/jamainternmed.2013.2784.

13. Ananworanich J, Schuetz A, Vandergeeten C, et al. Impact of multi-targeted antiretroviral treatment on gut T cell depletion and HIV reservoir seeding during acute HIV infection. *PLoS One*. 2012;7(3):e33948. doi:10.1371/journal.pone.0033948.

14. Chun T-W, Stuyver L, Mizell SB, et al. Presence of an inducible HIV-1 latent reservoir during highly active antiretroviral therapy. *Proc Natl Acad Sci*. 1997;94(24):13193-13197. doi:10.1073/pnas.94.24.13193.

15. Finzi D, Blankson J, Siliciano JD, et al. Latent infection of CD4+ T cells provides a mechanism for lifelong persistence of HIV-1, even in patients on effective combination therapy. *Nat Med*. 1999;5(5):512-517. doi:10.1038/8394.

16. Deeks SG, Autran B, Berkhout B, et al. Towards an HIV cure: a global scientific strategy. *Nat Rev Immunol*. 2012;12(8):607-614. doi:10.1038/nri3262.

17. Shan L, Deng K, Shroff NS, et al. Stimulation of HIV-1-specific cytolytic T lymphocytes facilitates elimination of latent viral reservoir after virus reactivation. *Immunity*. 2012;36(3):491-501. doi:10.1016/j.immuni.2012.01.014.

18. Stone D, Kiem H-P, Jerome KR. Targeted gene disruption to cure HIV. *Curr Opin HIV AIDS*. 2013;8(3):217-223. doi:10.1097/COH.0b013e32835f736c.

19. Thompson S, Clarke AR, Pow AM, Hooper ML, Melton DW. Germ line transmission and expression of a corrected HPRT gene produced by gene targeting in embryonic stem cells. *Cell*. 1989;56(2):313-321. http://www.ncbi.nlm.nih.gov/pubmed/2912572. Accessed August 25, 2014.

20. Thomas KR, Capecchi MR. Site-directed mutagenesis by gene targeting in mouse embryo-derived stem cells. *Cell*. 1987;51(3):503-512. http://www.ncbi.nlm.nih.gov/pubmed/2822260. Accessed August 25, 2014.

21. Aubert M, Boyle NM, Stone D, et al. In vitro Inactivation of Latent HSV by Targeted Mutagenesis Using an HSV-specific Homing Endonuclease. *Mol Ther Nucleic Acids*. 2014;3:e146. doi:10.1038/mtna.2013.75.

22. Schleifman EB, Bindra R, Leif J, et al. Targeted disruption of the CCR5 gene in human hematopoietic stem cells stimulated by peptide nucleic acids. *Chem Biol*. 2011;18(9):1189-1198. doi:10.1016/j.chembiol.2011.07.010.

23. Yuan J, Wang J, Crain K, et al. Zinc-finger nuclease editing of human cxcr4 promotes HIV-1 CD4(+) T cell resistance and enrichment. *Mol Ther*. 2012;20(4):849-859. doi:10.1038/mt.2011.310.

24. Maier D a, Brennan AL, Jiang S, et al. Efficient clinical scale gene modification via zinc finger nuclease-targeted disruption of the HIV co-receptor CCR5. *Hum Gene Ther*. 2013;24(3):245-258. doi:10.1089/hum.2012.172.

25. Miller JC, Tan S, Qiao G, et al. A TALE nuclease architecture for efficient genome editing. *Nat Biotechnol*. 2011;29(2):143-148. doi:10.1038/nbt.1755.

26. Vogelstein B, Kinzler KW. Digital PCR. *Proc Natl Acad Sci U S A*. 1999;96(16):9236-9241. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=17763&tool=pmcentrez&rendertype=abstract. Accessed September 24, 2014.

27. Sykes PJ, Neoh SH, Brisco MJ, Hughes E, Condon J, Morley AA. Quantitation of targets for PCR by use of limiting dilution. *Biotechniques*. 1992;13(3):444-449. http://www.ncbi.nlm.nih.gov/pubmed/1389177. Accessed August 7, 2014.

28. Baker M. Digital PCR hits its stride. *Nat Methods*. 2012;9(6):541-544. doi:10.1038/nmeth.2027.

29. Hindson BJ, Ness KD, Masquelier DA, et al. High-throughput droplet digital PCR system for absolute quantitation of DNA copy number. *Anal Chem*. 2011;83(22):8604-8610. doi:10.1021/ac202028g.

30. Kiss MM, Ortoleva-Donnelly L, Beer NR, et al. High-throughput quantitative polymerase chain reaction in picoliter droplets. *Anal Chem*. 2008;80(23):8975-8981. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2771884&tool=pmcentrez&rendertype=abstract. Accessed August 26, 2014.

31. Dingle TC, Sedlak RH, Cook L, Jerome K. Tolernce of Droplet-Digital PCR vs Real-Time Quantitative PCR to Inhibitory Substances. *Clin Chem*. 2013;59(11):1670-1672. doi:10.1373/clinchem.2013.206771.

32. Sedlak RH, Jerome KR. Viral diagnostics in the era of digital polymerase chain reaction. *Diagn Microbiol Infect Dis*. 2013;75(1):1-4. doi:10.1016/j.diagmicrobio.2012.10.009.

33. Strain MC, Lada SM, Luong T, et al. Highly precise measurement of HIV DNA by droplet digital PCR. Wu Y, ed. *PLoS One*. 2013;8(4):e55943. doi:10.1371/journal.pone.0055943.

34. Hayden RT, Gu Z, Ingersoll J, et al. Comparison of droplet digital PCR to real-time PCR for quantitative detection of cytomegalovirus. *J Clin Microbiol*. 2013;51(2):540-546. doi:10.1128/JCM.02620-12.

35. Sedlak RH, Cook L, Cheng A, Magaret A, Jerome KR. Clinical utility of droplet digital PCR for human cytomegalovirus. *J Clin Microbiol*. 2014;52(8):2844-2848. doi:10.1128/JCM.00803-14.

36. Henrich TJ, Gallien S, Li JZ, Pereyra F, Kuritzkes DR. *Low-Level Detection and Quantitation of Cellular HIV-1 DNA and 2-LTR Circles Using Droplet Digital PCR*.; 2012. http://www.sciencedirect.com/science/article/pii/S0166093412002984. Accessed January 27, 2014.

37. Oxnard GR, Paweletz CP, Kuang Y, et al. Noninvasive detection of response and resistance in EGFR-mutant lung cancer using quantitative next-generation genotyping of cell-free plasma DNA. *Clin Cancer Res*. 2014. doi:10.1158/1078-0432.CCR-13-2482.

38. Deng L, Qi Z, Zou B, et al. Digital detection of multiple minority mutants in stool DNA for noninvasive colorectal cancer diagnosis. *Anal Chem*. 2012;84(13):5645-5652. doi:10.1021/ac3008016.

39. Evans MI, Wright DA, Pergament E, Cuckle HS, Nicolaides KH. Digital PCR for noninvasive detection of aneuploidy: power analysis equations for feasibility. *Fetal Diagn Ther*. 2012;31(4):244-247. doi:10.1159/000337544.

40. Hahn S, Jackson LG, Kolla V, Mahyuddin AP, Choolani M. Noninvasive prenatal diagnosis of fetal aneuploidies and Mendelian disorders: new innovative strategies. *Expert Rev Mol Diagn*. 2009;9(6):613-621. doi:10.1586/erm.09.43.

41. Pinheiro LB, Coleman VA, Hindson CM, et al. Evaluation of a droplet digital polymerase chain reaction format for DNA copy number quantification. *Anal Chem*. 2012;84(2):1003-1011. doi:10.1021/ac202578x.

42. Sedlak RH, Cook L, Huang M-L, et al. Identification of chromosomally integrated human herpesvirus 6 by droplet digital PCR. *Clin Chem*. 2014;60(5):765-772. doi:10.1373/clinchem.2013.217240.

43. Boissel S, Jarjour J, Astrakhan A, et al. megaTALs: a rare-cleaving nuclease architecture for therapeutic genome engineering. *Nucleic Acids Res*. 2014;42(4):2591-2601. doi:10.1093/nar/gkt1224.

44. Silva G, Poirot L, Galetto R, et al. Meganucleases and other tools for targeted genome engineering: perspectives and challenges for gene therapy. *Curr Gene Ther*. 2011;11(1):11-27. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3267165&tool=pmcentrez&rendertype=abstract. Accessed August 21, 2014.

45. Holkers M, Maggio I, Liu J, et al. Differential integrity of TALE nuclease genes following adenoviral and lentiviral vector gene transfer into human cells. *Nucleic Acids Res*. 2013;41(5):e63. doi:10.1093/nar/gks1446.

46. Schiffer JT, Aubert M, Weber ND, Mintzer E, Stone D, Jerome KR. Targeted DNA mutagenesis for the cure of chronic viral infections. *J Virol*. 2012;86(17):8920-8936. doi:10.1128/JVI.00052-12.

47. Roberts JD, Bebenek K, Kunkel TA. The accuracy of reverse transcriptase from HIV-1. *Science*. 1988;242(4882):1171-1173. http://www.ncbi.nlm.nih.gov/pubmed/2460925. Accessed August 13, 2014.

48. Pillay D, Green H, Matthias R, et al. Estimating HIV-1 drug resistance in antiretroviral-treated individuals in the United Kingdom. *J Infect Dis*. 2005;192(6):967-973. doi:10.1086/432763.

49. Richman DD, Morton SC, Wrin T, et al. The prevalence of antiretroviral drug resistance in the United States. *AIDS*. 2004;18(10):1393-1401. http://www.ncbi.nlm.nih.gov/pubmed/15199315. Accessed August 26, 2014.

50. Liégeois F, Vella C, Eymard-Duvernay S, et al. Virological failure rates and HIV-1 drug resistance patterns in patients on first-line antiretroviral treatment in semirural and rural Gabon. *J Int AIDS Soc*. 2012;15(2):17985. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=3510650&tool=pmcentrez&rendertype=abstract. Accessed March 24, 2015.

51. Gibson DG, Young L, Chuang R-Y, Venter JC, Hutchison CA, Smith HO. Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat Methods*. 2009;6(5):343-345. doi:10.1038/nmeth.1318.

52. Kustikova OS, Wahlers A, Kuhlcke K, et al. Dose finding with retroviral vectors: correlation of retroviral vector copy numbers in single cells with gene transfer efficiency in a cell population. *Blood*. 2003;102(12):3934-3937. doi:10.1182/blood-2003-05-1424.

53. Weber ND, Stone D, Sedlak RH, et al. AAV-mediated delivery of zinc finger nucleases targeting hepatitis B virus inhibits active replication. *PLoS One*. 2014;9(5):e97579. doi:10.1371/journal.pone.0097579.

54. Armbruster DA, Pry T. Limit of blank, limit of detection and limit of quantitation. *Clin Biochem Rev*. 2008;29 Suppl 1:S49-S52. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=2556583&tool=pmcentrez&rendertype=abstract. Accessed March 5, 2015.

55. Davidson CJ, Zeringer E, Champion KJ, et al. Improving the limit of detection for Sanger sequencing: A comparison of methodologies for KRAS variant detection. *Biotechniques*. 2012;2012(3):182-188. doi:10.2144/000113913.

56. Vouillot L, Thélie A, Pollet N. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays To Detect Mutations Triggered by Engineered Nucleases. *G3 (Bethesda)*. 2015;5(3):407-415. doi:10.1534/g3.114.015834.

57. Vouillot L, Thelie a., Pollet N. Comparison of T7E1 and Surveyor Mismatch Cleavage Assays To Detect Mutations Triggered by Engineered Nucleases. *G3: Genes|Genomes|Genetics*. 2015;33(0). doi:10.1534/g3.114.015834.

58. Nishizawa M, Hattori J, Shiino T, et al. Highly-sensitive allele-specific PCR testing identifies a greater prevalence of transmitted HIV drug resistance in Japan. *PLoS One*. 2013;8(12):e83150. doi:10.1371/journal.pone.0083150.

59. Cousins MM, Swan D, Magaret CA, Hoover DR, Eshleman SH. Analysis of HIV using a high resolution melting (HRM) diversity assay: automation of HRM data analysis enhances the utility of the assay for analysis of HIV incidence. *PLoS One*. 2012;7(12):e51359. doi:10.1371/journal.pone.0051359.

60. Towler WI, James MM, Ray SC, et al. Analysis of HIV diversity using a high-resolution melting assay. *AIDS Res Hum Retroviruses*. 2010;26(8):913-918. doi:10.1089/aid.2009.0259.

61. James MM, Wang L, Donnell D, et al. Use of a high resolution melting assay to analyze HIV diversity in HIV-infected Ugandan children. *Pediatr Infect Dis J*. 2012;31(11):e222-e228. doi:10.1097/INF.0b013e3182678c3f.

62. Beck IA, Mahalanabis M, Pepper G, et al. Rapid and sensitive oligonucleotide ligation assay for detection of mutations in human immunodeficiency virus type 1 associated with high-level resistance to protease inhibitors. *J Clin Microbiol*. 2002;40(4):1413-1419. http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=140364&tool=pmcentrez&rendertype=abstract. Accessed June 25, 2015.

63. Beck IA, Crowell C, Kittoe R, et al. Optimization of the oligonucleotide ligation assay, a rapid and inexpensive test for detection of HIV-1 drug resistance mutations, for non-North American variants. *J Acquir Immune Defic Syndr*. 2008;48(4):418-427. doi:10.1097/QAI.0b013e31817ed7d7.