AIDS RESEARCH AND HUMAN RETROVIRUSES
Volume 31, Number 6, 2015
© Mary Ann Liebert, Inc.
DOI: 10.1089/aid.2014.0031

Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain

Jessica R. Keys,[1] Shuntai Zhou,[2] Jeffrey A. Anderson,[2,*] Joseph J. Eron, Jr.,[2]
Lauren A. Rackoff,[2] Cassandra Jabara,[3,†] and Ronald Swanstrom[4]

Abstract

Sequencing of a bulk polymerase chain reaction (PCR) product to identify drug resistance mutations informs antiretroviral therapy selection but has limited sensitivity for minority variants. Alternatively, deep sequencing is capable of detecting minority variants but is subject to sequencing errors and PCR resampling due to low input templates. We screened for resistance mutations among 184 HIV-1-infected, therapy-naive subjects using the 454 sequencing platform to sequence two amplicons spanning HIV-1 reverse transcriptase codons 34–245. Samples from 19 subjects were also analyzed using the MiSeq sequencing platform for comparison. Errors and PCR resampling were addressed by tagging each HIV-1 RNA template copy (i.e., cDNA) with a unique sequence tag (Primer ID), allowing a consensus sequence to be constructed for each original template from resampled sequences. In control reactions, Primer ID reduced 454 and MiSeq errors from 71 to 2.6 and from 24 to 1.2 errors/10,000 nucleotides, respectively. MiSeq also allowed accurate sequencing of codon 65, an important drug resistance position embedded in a homopolymeric run that is poorly resolved by the 454 platform. Excluding homopolymeric positions, 14% of subjects had evidence of ≥ 1 resistance mutation among Primer ID consensus sequences, compared to 2.7% by bulk population sequencing. When calls were restricted to mutations that appeared twice among consensus sequence populations, 6% of subjects had detectable resistance mutations. The use of Primer ID revealed 5–15% template utilization on average, limiting the depth of deep sequencing sampling and revealing sampling variation due to low template utilization. Primer ID addresses important limitations of deep sequencing and produces less biased estimates of low-level resistance mutations in the viral population.

Introduction

Combination antiretroviral therapy continues to improve patient outcomes as better treatment options are developed.[1–3] Advances may be offset by the development of resistance and cross-resistance among subjects failing multiple regimens,[4,5] which may also be transmitted to susceptible partners.[6,7] Since transmitted drug resistance may compromise patient response to first line combination therapy,[8–10] genotypic resistance testing is routinely recommended before therapy initiation.[11] The utility of pretherapy testing may be limited by minority HIV-1 variants present in $<20\%$ of the viral population that are not reliably detected by standard sequencing,[12] and that may jeopardize virologic response.[13–17]

The prevalence of minority pretherapy drug resistance mutations varies, with estimates based on highly sensitive research assays often double those reported using standard sequence analysis.[14–19] Some variation may be related to methods for measuring low abundance resistance mutations. For example, allele-specific polymerase chain reaction (PCR) detects mutations that make up $\geq 0.01\%$ of the viral population,[15] which would require at least 30,000 input templates for reasonable sampling; however, prevalence estimates are based on a few, predetermined mutations, and estimates of resistance mutations in an individual may be biased by differential amplification. Alternatively, single genome sequencing allows interrogation of entire viral genomes diluted to a single viral sequence, bypassing some

[1] Department of Epidemiology, University of North Carolina, Chapel Hill, North Carolina.
[2] Division of Infectious Diseases, Department of Medicine, University of North Carolina, Chapel Hill, North Carolina.
[3] Department of Biology, University of North Carolina, Chapel Hill, North Carolina.
[4] Department of Biochemistry and Biophysics, University of North Carolina, Chapel Hill, North Carolina.
*Current affiliation: Discovery Medicine and Clinical Pharmacology, Bristol-Myers Squibb, Pennington, New Jersey.
†Current affiliation: Assay Research and Development, Abbott Molecular, Des Plaines, Illinois.

amplification errors, but this labor-intensive method generally achieves low sensitivity due to limited sampling depth.[12] In contrast, ultradeep sequencing involves massively parallel sequencing of amplified viral sequences, producing upward of 10^6 sequences and theoretically enabling detection of 1% variants or less[20]; however, this method frequently generates errors during amplification and sequencing, making it difficult to distinguish true minority variants from sequencing errors.[21] In addition, resequencing of a smaller number of viral genomes after PCR amplification (PCR resampling) gives overestimates of the true sampling depth.

To address errors associated with deep sequencing, threshold cut-offs based on estimated error rates from known sequences have been established.[22] Cut-offs do not account for errors that may be introduced during the PCR step, such as biased amplification and nucleotide misincorporation,[23,24] nor do they address PCR resampling. Here, an alternative strategy engineered into the cDNA synthesis step (Primer ID) was used to circumvent the need for statistically defined cut-offs[25] allowing (1) definition of a background error rate for two deep sequencing platforms; (2) estimation of the prevalence of preexisting reverse transcriptase inhibitor (RTI) resistance mutations among chronically infected subjects; and (3) comparison of prevalence estimates from standard methods to those obtained by deep sequencing.

Materials and Methods

Study population

Study subjects were previously enrolled in the University of North Carolina Center for AIDS Research HIV Clinical Cohort Study (UCHCC).[26] UCHCC is an ongoing, clinical cohort enrolling HIV-infected adults receiving care at UNC. UCHCC maintains an electronic database of patient information and houses a repository of plasma samples obtained during routine care. UCHCC subjects were eligible if they (1) initiated therapy after December 31,1999 with two or more nucleoside/tide reverse transcriptase inhibitors (NRTI) plus one nonnucleoside reverse transcriptase inhibitor (NNRTI), or three or more NRTI; and (2) had at least one reported pretherapy HIV-1 RNA level. This study was reviewed and approved by the University of North Carolina Institutional Review Board.

HIV-1 sequencing

For most subjects ($N=141/184$), bulk sequencing analyses were obtained using commercial HIV-1 GenoSure (Plus) assays (LabCorp, Research Triangle Park, NC). If no bulk sequence analysis was available ($N=43/184$), we attempted in-house sequencing of HIV-1 reverse transcriptase (RT) codons 34–245 using the ABI Prism BigDye Version 1.1 Terminal Cycle Sequencing (Life Technologies, Carlsbad, CA). To check for evidence of cross-contamination, sequences were aligned by Clustal W version 2[27] and inspected by constructing neighbor-joining phylogenetic trees evaluated with 1,000 bootstrap replicates.[28]

Sample amplicon libraries were generated using previously described methods.[25] Samples with $<4.5 \log_{10}$ HIV-1 RNA copies/ml were centrifuged to concentrate the virus particles prior to RNA extraction (QIAamp viral RNA extraction kit, Qiagen, Hilden, Germany). One-third of the RNA was used in separate cDNA synthesis reactions targeting two regions of HIV-1 RT, HXB2 nucleotides 2648–2964 and 2965–3257 (RT codons 34–139 and 139–245)[17] using the primers listed in Supplementary Table S1 (Supplementary Data are available online at www.liebertpub.com/aid). The cDNA primers included a barcode to allow pooling of samples during the deep sequencing step and a randomized sequence tag of eight nucleotides (Primer ID) to allow identification of each individual template in the subsequent sequence analysis (Fig. 1A). Purified cDNA[25] was used for seminested PCR (Fig. 1B) using Phusion Hot Start II High Fidelity DNA polymerase (Thermo Fisher Scientific, Waltham, MA); annealing temperatures were 67°C and 63°C for RT fragments 1 and 2, respectively (Supplementary Table S1). Input cDNA was initially estimated based on the assumption that all RNA templates (500 copies) were copied into cDNA.

Amplified DNAs were pooled in equimolar concentrations and gel purified (QIAquick gel purification kit, Qiagen). Pools were submitted for sequencing on the 454 GS FLX sequencing platform with XLR80 titanium reagents (Roche, Indianapolis, IN). To compare 454 and MiSeq nucleotide calls, 19 amplicons were also sequenced using the 150-bp paired-end sequencing protocol (HIV-1 RT codons 34–74 and 111–139, HXB2 nucleotides 2648–2770 and 2878–2964) on the Illumina MiSeq sequencing platform (San Diego, CA). Illumina MiSeq adapters were added during the initial round of PCR.

Plasmid controls

The entirety of HIV-1 RT (HXB2 nucleotides 2550–3515), derived from a clinical sample, was cloned into vector pcDNA3.1 (Life Technologies). Plasmids were linearized with *Bam*HI, purified using the Minelute PCR purification kit (Qiagen), quantified by UV spectrophotometry using a Nanodrop 1000 (Thermo Fisher Scientific), and serially diluted to 3,000, 10,000, 30,000, and 300,000 copies. Plasmid DNA dilutions were denatured at 95°C for 5 min, cooled, and tagged with distinct cDNA primers, including Primer ID, during a single round of DNA synthesis with Platinum Taq (Life Technologies). Excess cDNA primers were removed[25] and samples were amplified by nested PCR using a protocol identical to samples with primers listed in Supplementary Table S1, pooled in equimolar concentration, and gel purified (Qiagen). Pools were sequenced over HIV-1 RT codons 34–139 using the Roche 454 Junior (HXB2 nucleotides 2648–2964) and Illumina MiSeq sequencing platforms (HXB2 nucleotides 2648–2868 and 2782–2964).

Sequence analysis

Deep sequencing data were processed using a custom pipeline of computer programs,[25] which is available by request along with a sample dataset. Briefly, sequence length distributions were inspected, short reads and reads with ambiguous base calls were discarded, and the remaining sequences were compared to HXB2 *pol* for orientation and location. Sequences with an invalid Primer ID or barcode were discarded, and filtered sequences were partitioned first by barcode (sample) and then Primer ID (viral template). Sequences with a Primer ID that occurred less than three times within a sample were discarded (less than five times for MiSeq), and consensus sequences were generated from

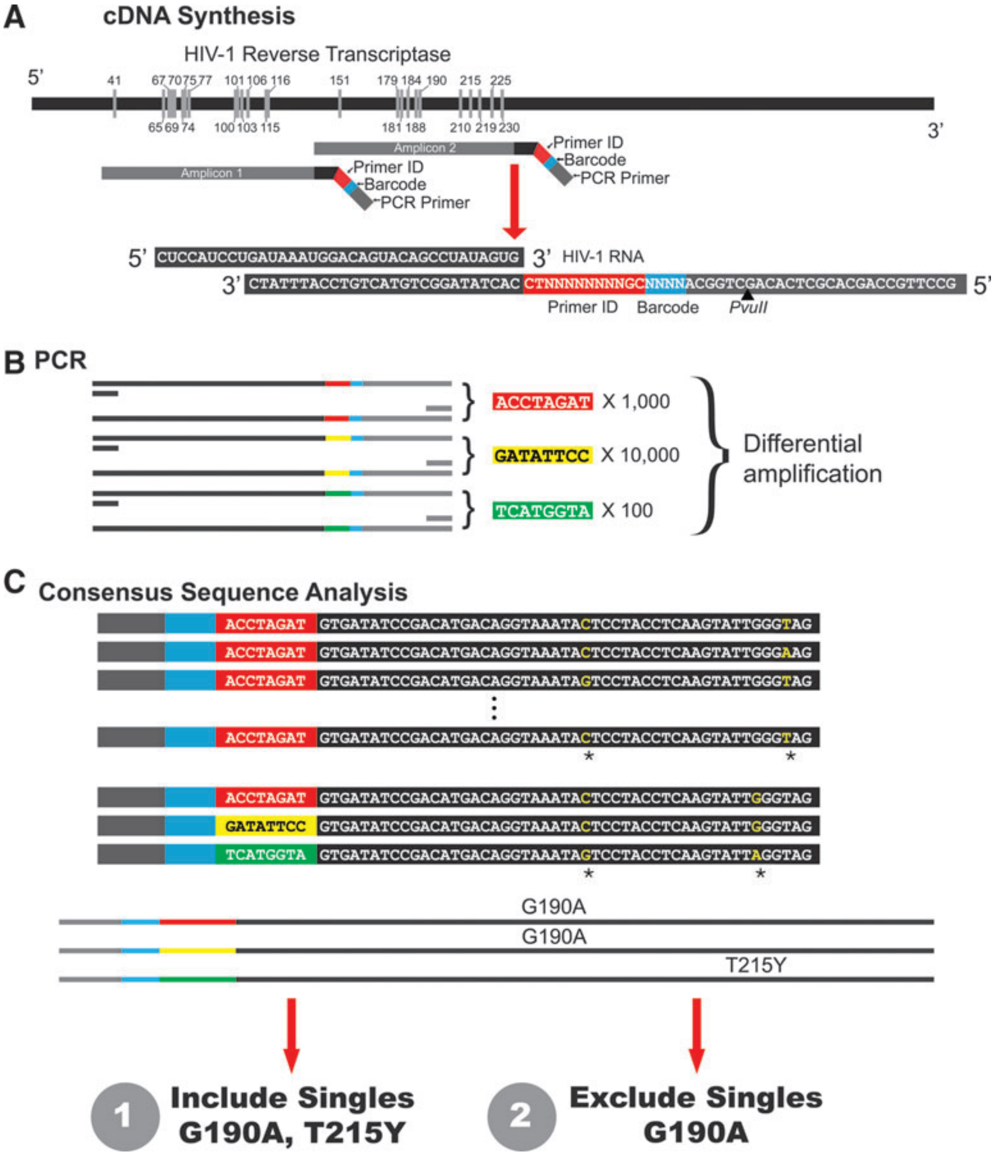

A
cDNA Synthesis
HIV-1 Reverse Transcriptase
5'
41
67 70 75 77
101 106 116
151
179 184 190
215
225
65 69 74
100 103 115
181 188
210 219 230
3'
Amplicon 2
Primer ID
Barcode
PCR Primer
Amplicon 1
Primer ID
Barcode
PCR Primer
5' CUCCAUCCUGAUAAAUGGACAGUACAGCCUAUAGUG 3' HIV-1 RNA
3' CTATTTACCTGTCATGTCGGATATCAC CTNNNNNNNNGC NNNN ACGGTCGACACTCGCACGACCGTTCCG 5'
Primer ID
Barcode
PvuII
B PCR
ACCTAGAT X 1,000
GATATTCC X 10,000
TCATGGTA X 100
Differential amplification
C Consensus Sequence Analysis
ACCTAGAT GTGATATCCGACATGACAGGTAAATACTCCTACCTCAAGTATTGGGTAG
ACCTAGAT GTGATATCCGACATGACAGGTAAATACTCCTACCTCAAGTATTGGGAAG
ACCTAGAT GTGATATCCGACATGACAGGTAAATAGTCCTACCTCAAGTATTGGGTAG
ACCTAGAT GTGATATCCGACATGACAGGTAAATACTCCTACCTCAAGTATTGGGTAG
*
*
ACCTAGAT GTGATATCCGACATGACAGGTAAATACTCCTACCTCAAGTATTGGGTAG
GATATTCC GTGATATCCGACATGACAGGTAAATACTCCTACCTCAAGTATTGGGTAG
TCATGGTA GTGATATCCGACATGACAGGTAAATAGTCCTACCTCAAGTATTAGGTAG
*
*
G190A
G190A
T215Y
1 Include Singles G190A, T215Y
2 Exclude Singles G190A

FIG. 1. Primer ID method to estimate the HIV-1 population. **(A)** A unique Primer ID sequence is incorporated into each viral genome along with a sample-specific barcode (*blue*) during cDNA synthesis. For each clinical subject sample, two independent cDNA synthesis reactions were set up to query most of HIV-1 reverse transcriptase (RT) codons 34-245, shown **(A)** as Amplicon 1 and 2. Amplicon 2 is enlarged to illustrate the details of the cDNA primer. **(B)** Differential amplification may occur during polymerase chain reaction (PCR) so that the probability of amplification is not equally distributed across individual viral genomes. **(C)** Illustration of how Primer ID is applied to correct errors that accumulated over a deep sequencing run. Within each barcoded (*blue*) sample, a single majority-rules consensus sequence is generated from three or more sequences with the same Primer ID (*red*, *green*, *yellow*). Collectively, consensus sequences reflect the actual number of viral genomes tagged during cDNA synthesis rather than what is best amplified. Method (2) is more conservative than (1) since single occurrences of resistance mutations on Primer ID consensus sequences are also excluded as error.

sequences with the same Primer ID. Only 1.5% of sequences were discarded because their Primer ID occurred on less than three sequences within a sample. A larger number of identical Primer ID reads was used to build the consensus sequence for data from the MiSeq platform because of the larger number of reads available and with the goal of obtaining a more reliable consensus sequence.

Drug resistance was defined using an updated list of surveillance drug resistance mutations to exclude polymorphisms that may not contribute to a resistance phenotype.[29] Surveillance drug resistance mutations detected by deep sequencing were quantified using two conditions (Fig. 1C): (1) barcode- and Primer ID-defined consensus sequences and (2) barcode- and Primer ID-defined consensus sequences, but only including resistance mutations that occurred in more than one consensus sequence within a sample (i.e., associated with two or more different Primer ID consensus sequences). The relative abundance of individual resistance mutations per sample was calculated by dividing the number of sequences with at least one resistance mutation by the total number of consensus sequences obtained for the sample.

Deep sequencing-detected resistance mutations were considered in the context of homopolymeric regions, which are error hotspots for the 454 sequencing platform.[21,30] We defined homopolymeric regions as four or more consecutive, identical nucleotides plus the two flanking nucleotides. We used the 2004 HIV-1 subtype B consensus sequence to define homopolymer-associated positions for clinical subject samples (available from the Los Alamos HIV Database at www.hiv.lanl.gov), and we defined homopolymeric regions directly for the subtype C plasmid control. In the subtype C control sequence, 75 homopolymer-associated positions were identified within the 317 nucleotide sequence spanning HXB2 2648–2964. In the 2004 subtype B consensus sequence, 61 and 45 homopolymer-associated positions were identified within this sequence and the downstream 265 nucleotide sequence spanning HXB2 2993–3257. Since mutations within HIV-1 RT codons 65, 67, 74, 100, 101, 103, 115, 116, and 219 were within regions influenced by homopolymeric tracts, they were excluded from overall prevalence estimates. However, some of these positions were included in the comparative analysis with the MiSeq data.

Statistical analysis

Standard errors and 95% confidence intervals (CI) for plasmid control error rates (errors per 10,000 nucleotides in consensus sequences) were calculated across all samples using clustered sandwich estimators[31] and the Poisson distribution. Standard errors and 95% CI for proportions were estimated using the binomial distribution. Sequencing depth, or the number of sequences required to observe *x*% viral variant with 95% confidence, was estimated using a power analysis. Distributions of categorical variables were compared using Pearson's χ^2 test and median values of continuously distributed variables were compared using the Kruskal–Wallis test. Statistical analyses were conducted in SAS version 9.3 (SAS Institute, Cary, NC).

Results

Quantifying deep sequencing error

The goal of this study was to examine low-level resistance to RT inhibitors encoded within the RT coding domain. Control amplicons were designed to include several important resistance positions: RT codons 34–139 or HXB2 nucleotides 2648–2964. This region was selected for control experiments since it is richer in homopolymers compared to the downstream amplicon spanning HIV-1 RT codons 138–245 (42% of nucleotides near homopolymer positions in amplicon 1 versus 34% in amplicon 2), and since many clinically important resistance mutations are located in this sequence, including K103N and K65R.

First, we established a residual error rate for both the 454 and Miseq platforms using plasmid controls to evaluate our ability to interpret rare variants using either the raw sequences or sequences corrected by Primer ID, which was used as the primer in the first round of DNA synthesis. We used Taq DNA polymerase rather than RT in the first round of synthesis because of low template utilization by RT when starting with a DNA template; also, we used a DNA template rather than an RNA template to avoid misincorporation during the synthesis of RNA *in vitro*. A total of 112,108 reads of the 317 nucleotide amplicon obtained using the 454 platform were collapsed into 2,893 Primer ID consensus sequences across all samples. The overall error rate using raw sequences was 71/10,000 nucleotides (95% CI: 70–72), which was reduced to 2.6/10,000 nucleotides (95% CI: 2.2–3.2) using the Primer ID/consensus sequence approach (Supplementary Table S2). Over 75 homopolymeric positions, 6.0 (95%CI: 4.8–7.4) miscalls were observed every 10,000 nucleotides using Primer ID; excluding homopolymeric regions reduced the error to 1.6/10,000 (95% CI: 1.3–2.0) nucleotides. Errors were substitutions (76%), deletions (22%), or insertions (1.7%). Error frequency is compared for each position queried in Fig. 2A.

We were especially interested in the potential impact of homopolymeric error in the region of the K65 codon, the position of an important resistance mutation for tenofovir. In most subtype B isolates the lysine codon at this position is AAA and part of a longer homopolymeric region. However, some subtype B and most subtype C isolates have the AAG codon embedded in this longer homopolymer, and misplacement of the G can create the appearance of the AGA codon at this position, which would be interpreted as a resistance mutation (arginine). In the control plasmid, which had a K65 AAG codon, only 59% of raw sequences had the correct sequence and 38% of the sequences were ''ATA AAA A-G AAA GAC.'' This was caused by the undercall of an A in the homopolymeric region in front of the AAG codon, thus shifting the G one position to the left and creating an AGA codon. Since it is not possible to know which A of the homopolymeric tract was undercounted, the presence of the AGA codon cannot be interpreted. A more difficult situation would be when there is an undercall in the upstream portion of the homopolymeric tract and an overcall in the downstream portion, which would shift the G and leave the reading frame intact, thus obscuring the effect of the homopolymeric errors and creating an erroneous call for a resistance mutation. This was a very rare event that occurred in 12/112,108 raw reads.

Alternatively, deep sequencing with the MiSeq platform is not susceptible to homopolymeric miscalls because it does not rely on the linearity of the relationship between the number of consecutive identical nucleotides and signal peak. Using the MiSeq sequencing platform, 123,822 raw reads were sampled from 678,702 total paired-end reads of the 317 nucleotide control amplicon that passed Illumina quality filters. From these, 2,710 Primer ID consensus sequences were generated. Using the MiSeq platform to sequence the same set of controls yielded an error rate of 1.2/10,000 bases (95% CI: 0.59–2.4) compared to 24/10,000 bases (95% CI: 18–32) using raw sequence data (Supplementary Table S2). All errors were substitutions, and no difference was observed within homopolymeric regions (1.1 errors/10,000 nucleotides, 95% CI: 0.51–2.5). However, higher error rates were observed in the downstream compared to the upstream paired-end sequence (rate ratio: 2.9, 95% CI: 1.7–5.0). Within the downstream sequence, errors increased over the run 2.6 times per 100 nucleotides sequenced (95% CI: 1.3–5.0). Positional errors associated with the MiSeq platform are shown in Fig. 2B.

Prevalence of pretherapy drug resistance mutations

Of 331 eligible subjects in the UCHCC, 184 (56%) had an archived pretherapy sample. Most were chronically infected,

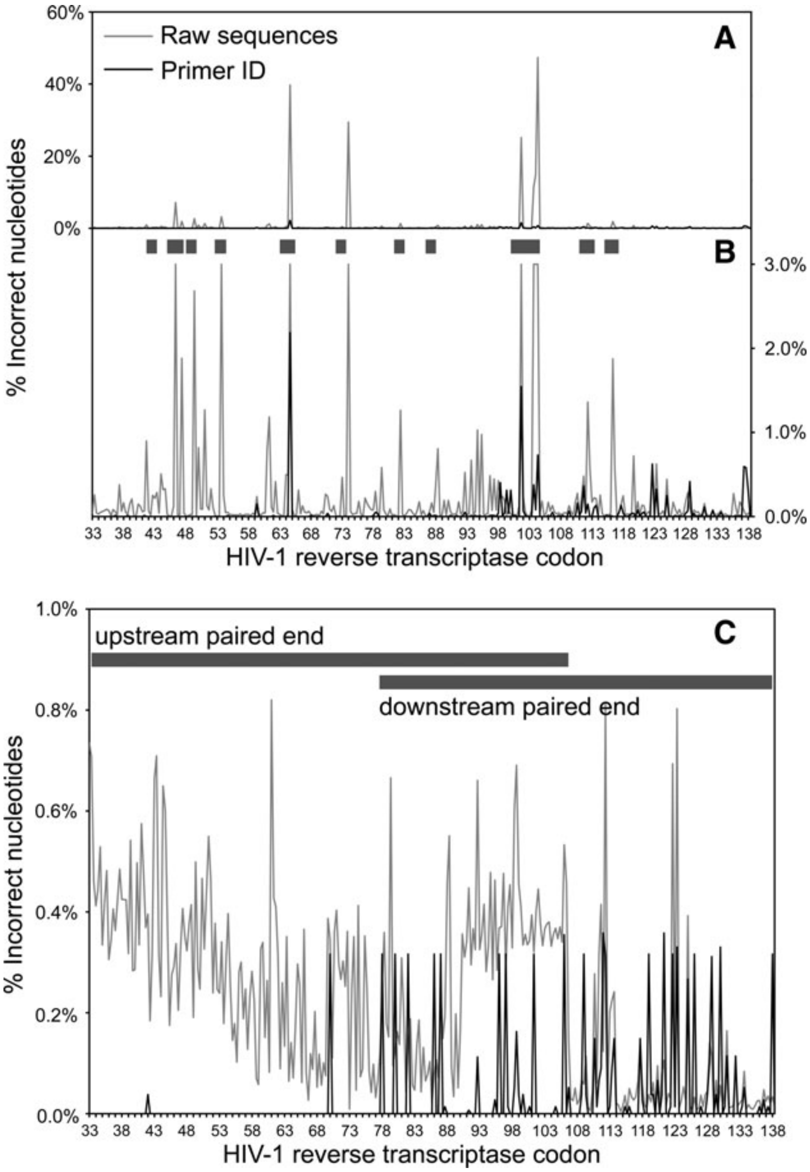


FIG. 2. Percent incorrect nucleotide calls for each queried position of control sequence. **(A, B)** Positional errors for sequences read by the 454 deep sequencing platform. **(B)** This is graphed using a smaller *y*-axis scale to visualize very low error frequency observed after Primer ID correction of raw sequences. Homopolymeric tracts are highlighted along the *x*-axis using *dark gray bars*. **(C)** Positional errors for sequences read using the Illumina MiSeq deep sequencing platform. The read length of each MiSeq paired end is highlighted along the *x*-axis using *dark gray bars*. Percent incorrect nucleotides was calculated for each nucleotide position by comparing raw sequences (*gray line*) or Primer ID consensus sequences (*black line*) to known plasmid control sequences. Error frequency was calculated for each platform across all dilutions submitted for deep sequencing.

with median 254 [interquartile range (IQR): 95–398] CD4$^+$ T cells/μl and 4.8 (IQR: 4.2–5.3) $\log_{10}$ HIV-1 RNA copies/ml prior to therapy (Table 1). The sequence analysis for these 184 subjects was based on the identification of resistance mutations to any RTI, or specifically to an NRTI or an NNRTI, with the resistance mutations defined by the 2009 list of surveillance drug resistance mutations.[29] We excluded homopolymer-associated positions from these prevalence estimates to allow for later comparison to the 454 platform. Based on sequencing of a bulk PCR product, 2.7% of subjects ($N=5/184$, 95% CI: 0.89–6.2%) had an RTI resistance-associated mutation. NRTI-associated resistance mutations were the most common, being present in 2.2% of subjects ($N=4/184$, 95% CI: 0.60–5.5%), while 1.1% of subjects ($N=2/184$, 95% CI: 0.13–3.4%) had an NNRTI resistance-associated mutation. The RTI resistance mutations detected by bulk sequencing in these five subjects are shown in Fig. 3A.

Across 184 subjects, $>10^6$ raw sequences were generated using the 454 sequencing platform, 73% ($N=746{,}809$) of which were >300 nucleotides long. Excluding mutations below 1% abundance in the raw reads, 21% of subjects ($N=38/184$, 95% CI: 15–27%) had any RT inhibitor resistance mutations, including 18% ($N=34/184$, 95% CI: 13–25%) and 7.1% ($N=13/184$, 95% CI: 3.8–12%) with an NRTI or NNRTI resistance mutation (Supplement Table S3). However, these estimates overlook the effects of allelic bias during PCR amplification, PCR resampling, and potential hotspots for error incorporation.

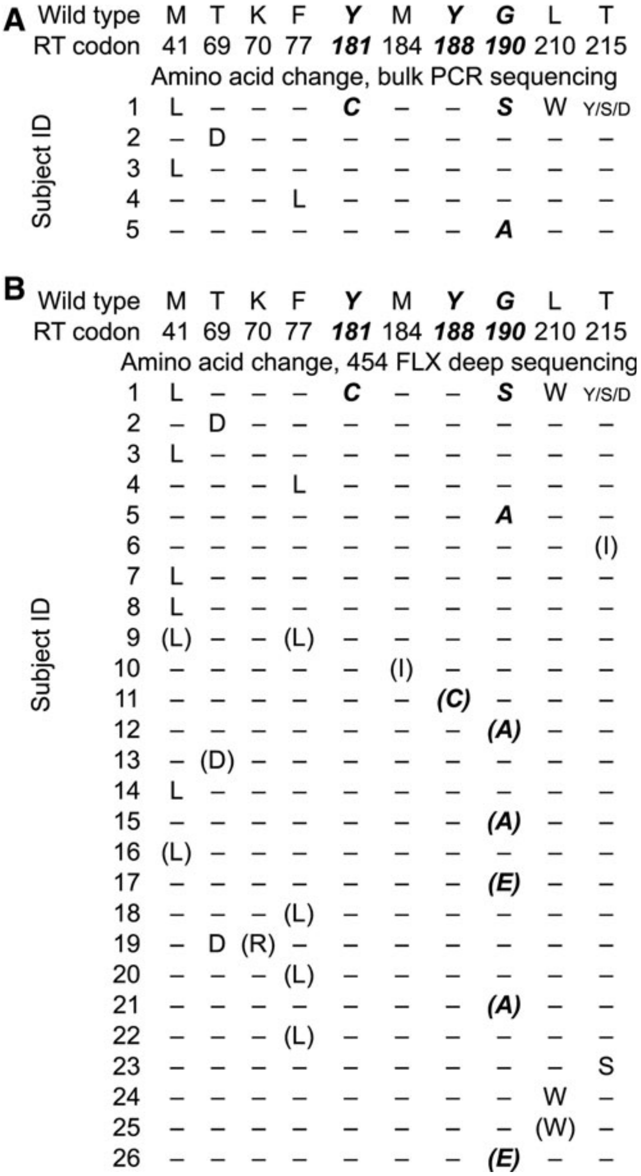
We next used the Primer IDs to form consensus sequences from the reads that represented PCR resampling. Across both amplicons, a median 1,475 (IQR: 598–2,471) raw sequences per subject were collapsed into a median 41 (IQR: 18–76) consensus sequences per subject, corresponding to an average sequencing depth for reliable detection of about 7% (IQR: 4–17%). The large reduction in usable reads from the raw reads to the consensus sequences is a function of removing PCR resampling with Primer ID tagging to reveal the actual number of templates sampled. We observed that only 5–15% of the RNA templates added to the cDNA reactions

Table 1. Baseline Characteristics of Clinical Subjects

	All eligible	*Sample available*
Characteristic	N=*331*	N=*184*
Gender, *n* (%)		
Female	77 (23%)	42 (23%)
Male	254 (77%)	142 (77%)
Race, *n* (%)		
Black	181 (55%)	98 (53%)
White	92 (28%)	51 (28%)
Other	58 (17%)	35 (19%)
Age, median (IQR)[a]	38 (31–46)	38 (31–47)
HIV risk group		
MSM, *n* (%)	144 (44%)	80 (43%)
IDU, *n* (%)	29 (8.8%)	15 (8.1%)
Heterosexual, *n* (%)	196 (59%)	110 (60%)
Year of first therapy, *n* (%)		
1999–2001	104 (31%)	38 (21%)
2002–2004	99 (30%)	64 (35%)
2005–2007	83 (25%)	54 (29%)
>2007	45 (14%)	28 (15%)
First regimen, *n* (%)		
NRTI only	45 (14%)	22 (12%)
NVP	22 (6.7%)	9 (4.9%)
EFV	264 (80%)	153 (83%)
HIV-1 RNA $\log_{10}$ copies/ml, median (IQR)	4.8 (4.3–5.4)	4.8 (4.2–5.3)
$CD4^+$T cells/μl, median (IQR)	205 (54–357)	254 (95–398)

[a] Age is calculated using the date of antiretroviral therapy initiation.

IQR, interquartile range; MSM, men who have sex with men; IDU, injection drug use; NRTI, nucleoside reverse transcriptase inhibitor; NVP, nevirapine; EFV, efavirenz.



A

Wild type	M	T	K	F	***Y***	M	***Y***	***G***	L	T
RT codon	41	69	70	77	***181***	184	***188***	***190***	210	215
Subject ID	Amino acid change, bulk PCR sequencing									
1	L	–	–	–	***C***	–	–	***S***	W	Y/S/D
2	–	D	–	–	–	–	–	–	–	–
3	L	–	–	–	–	–	–	–	–	–
4	–	–	–	L	–	–	–	–	–	–
5	–	–	–	–	–	–	–	***A***	–	–

B

Wild type	M	T	K	F	***Y***	M	***Y***	***G***	L	T
RT codon	41	69	70	77	***181***	184	***188***	***190***	210	215
Subject ID	Amino acid change, 454 FLX deep sequencing									
1	L	–	–	–	***C***	–	–	***S***	W	Y/S/D
2	–	D	–	–	–	–	–	–	–	–
3	L	–	–	–	–	–	–	–	–	–
4	–	–	–	L	–	–	–	–	–	–
5	–	–	–	–	–	–	–	***A***	–	–
6	–	–	–	–	–	–	–	–	–	(I)
7	L	–	–	–	–	–	–	–	–	–
8	L	–	–	–	–	–	–	–	–	–
9	(L)	–	–	(L)	–	–	–	–	–	–
10	–	–	–	–	–	(I)	–	–	–	–
11	–	–	–	–	–	–	***(C)***	–	–	–
12	–	–	–	–	–	–	–	***(A)***	–	–
13	–	(D)	–	–	–	–	–	–	–	–
14	L	–	–	–	–	–	–	–	–	–
15	–	–	–	–	–	–	–	***(A)***	–	–
16	(L)	–	–	–	–	–	–	–	–	–
17	–	–	–	–	–	–	–	***(E)***	–	–
18	–	–	–	(L)	–	–	–	–	–	–
19	–	D	(R)	–	–	–	–	–	–	–
20	–	–	–	(L)	–	–	–	–	–	–
21	–	–	–	–	–	–	–	***(A)***	–	–
22	–	–	–	(L)	–	–	–	–	–	–
23	–	–	–	–	–	–	–	–	–	S
24	–	–	–	–	–	–	–	–	W	–
25	–	–	–	–	–	–	–	–	(W)	–
26	–	–	–	–	–	–	–	***(E)***	–	–

FIG. 3. Subjects with preexisting RT inhibitor resistance mutations. **(A)** The RT inhibitor resistance genotype for each of five subjects with mutations detected using standard sequence analysis. **(B)** The RT inhibitor resistance genotype for 26 subjects with resistance mutations detected using the 454 FLX deep sequencing platform, corrected using Primer ID. Mutations associated with a single Primer ID consensus sequence within a subject sample are shown in *parentheses*. RT codons and mutations associated with nonnucleoside reverse transcriptase inhibitor (NNRTI) resistance are highlighted in *bold italic type*, while RT codons and mutations associated with nucleoside reverse transcriptase inhibitor (NRT)I resistance are shown in *standard type*. Only RT codons outside of homopolymeric influence were included in this analysis. Lack of a particular substitution associated with RT inhibitor resistance is indicated by a *dash*.

resulted in consensus sequences, indicating inefficient cDNA priming and/or extension, or inefficient inclusion of cDNA products into the PCR.

In our first analysis using Primer ID, a resistance mutation was considered if it appeared in any consensus sequence created using Primer ID, even if it appeared in a single consensus sequence (Fig. 1C, method 1). A total of 14% (*N*=26/184, 95% CI: 9.4–20%) of subjects had RTI resistance-associated mutations among Primer ID consensus sequences using the 454 platform, including 11% (*N*=20/184, 95% CI: 6.8–16%) of subjects with NRTI resistance mutations and 4.9% (*N*=9/184, 95% CI: 2.3–9.1%) of subjects with NNRTI resistance mutations. All of the RTI resistance mutations observed by bulk sequencing were also observed in the Primer ID consensus sequences (Fig. 3B). Conversely, using Primer ID consensus sequences rather than raw sequences resulted in a 33% reduction in the number of subjects where a resistance mutation was observed (21% versus 14%) after using a conservative (but arbitrary) 1% cut-off for mutations in the raw reads.

The frequency of single mutations of any type in the data set of Primer ID consensus sequences was four times higher than expected given the error rate determined using the plasmid sequences. Thus, in most cases the call of a resistance mutation based on a single observation was likely accurate. However, in a second, more conservative analysis, only those resistance mutations that appeared in at least two consensus sequences were counted (Fig. 1C, method 2).

When only multiple (i.e., two or more) within-subject observations of a specific resistance mutation were included, the prevalence of RTI resistance mutations among these 184 subjects was 6.0% ($N=11/184$, 95% CI: 3.0–10%), representing six additional subjects over the five who were also identified using bulk sequence analysis (Fig. 3B).

The preceding analysis did not include the homopolymeric regions, and we carried out a separate analysis to see what influence they would have on calls of drug resistance mutations. We found that only four (2%) out of 184 subjects sequenced using 454 had a predominant ''AAG'' (a wild-type Lys codon) at RT codon 65, and 11 (6%) subjects had an ''AAG'' at RT codon 65 as a minority variant with abundance ranging from 1% to 24%. No evidence of K65R was found. If other homopolymeric positions were included, an additional 24 subjects would have been classified as RTI resistant using the 454 data with consensus sequences, raising the overall prevalence to 27% ($N=50/184$). Some calls at homopolymeric positions were also seen by bulk sequence analysis (in three subjects) and were unlikely due to homopolymeric error given their high abundance. In contrast, homopolymer-associated resistance mutations detected solely by deep sequencing ranged in frequency from 0.35% to 12.5% and most appeared once within a sample. Assuming these single occurrences were miscalls due to homopolymeric error, only six additional subjects would be classified as having preexisting RTI resistance (based on the mutation being on more than one consensus sequence). Thus, if homopolymer-associated positions were included, prevalence estimates would increase to 9% ($N=17/184$).

Relative abundance of resistance mutations within viral populations

The use of Primer ID allows an assessment of the number of viral genomes that were actually sampled from a subject, thus allowing an assessment of both sequencing depth and the relative abundance of detected mutations from a specific clinical sample. There were six subjects who had resistance mutations that were detected in multiple Primer ID consensus sequences but not detected by bulk sequencing. The median abundance of these mutations within the viral population in each person, determined using the maximum proportion of sequences in which the mutation was observed, was 1.8% (IQR: 1.2–2.8%). These subjects had a higher than average number of consensus sequences compared to the entire population, with a median 159 consensus sequences (IQR: 106.5–235.5) per subject across both amplicons. There were an additional 15 subjects who had a resistance mutation present in one Primer ID consensus sequence and the median abundance of these mutations was 1.7% (IQR: 0.67–1.1%), detected among a median 73 consensus sequences per subject (IQR: 44–149). However, although these groups of subjects had more consensus sequences on average, the estimate of abundance is significantly limited given the low number of observations of each mutation, and the ability to detect variants at even less abundance is limited by low template utilization, as revealed using the Primer ID. This phenomenon is highlighted in Fig. 4, where the majority of low-abundance resistance mutations were detected on a single Primer ID consensus sequence within a subject sample.

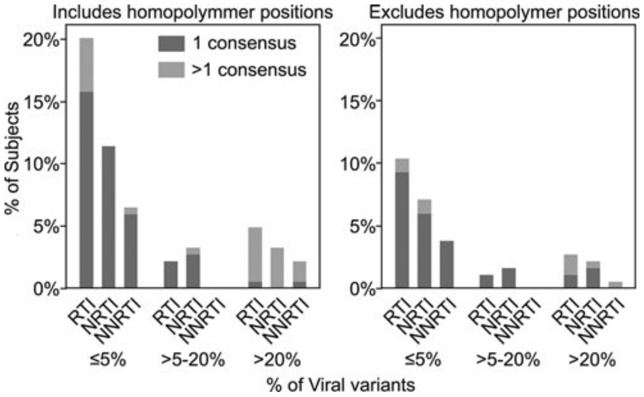


FIG. 4. Prevalence and relative abundance of preexisting RT inhibitor resistance mutations among clinical subjects ($N=184$). All estimates are derived from Primer ID corrected deep sequencing using the 454 FLX platform. The *right-hand panel* excludes RT positions near homopolymeric tracts, defined as four or more consecutive, identical nucleotides plus the two flanking nucleotides. The *left-hand panel* includes all RT inhibitor resistance mutations, even those near homopolymeric tracts. RT inhibitor resistance mutations were defined using an updated list of surveillance drug resistance mutations.[29]

Using Primer ID, few subjects with RTI resistance mutations had evidence of multiple resistance mutations. Of 21 subjects with minority drug resistance mutations, only two (10%) had more than one drug resistance mutation, each occurring on separate Primer ID consensus sequences at very low frequency. Of five subjects with a majority drug-resistant population, only one (20%) had multiple resistance mutations. This subject had multiple drug resistance mutations that were also revealed by bulk sequence analysis: Y181C, G190S, and L210W appeared with T215Y, T215S, or T215D among 79% ($N=26/29$), 10% ($N=3/29$), or 10% ($N=3/29$) of consensus sequences, respectively, while M41L was linked to homopolymer-associated L74V and K101E in 94% ($N=29/31$) of consensus sequences. Together, this suggests that this subject was initially infected with a variant carrying M41L, L74V, K101E, Y181C, G190S, L210W, and T215Y mutations, with the virus slowly reverting at codons 74 and 215.

Comparison of deep sequencing platforms in a clinical setting

Sequences spanning HIV-1 RT codons 34–74 and 111–139 (HXB2 nucleotides 2648–2770 and 2878–2964) were determined for 19 of 184 subjects using the Illumina MiSeq platform. Based on previous analyses of the data from the 454 FLX sequencing platform, we selected subjects who had the most consensus sequences constructed from ≥3 raw sequences sharing the same Primer ID [median 203 (IQR: 168–247) consensus sequences], indicating that these samples had the highest level of genomes incorporated into the cDNA/PCR step. Using the MiSeq platform, a median 273 (IQR: 192–583) consensus sequences were constructed from ≥5 raw sequences sharing a Primer ID [median 29,743 (IQR): 24,686–33,086 raw sequences]. For 17 of the 19 subjects, the number of consensus sequences generated using the MiSeq

Table 2. Percentage of K103N Detected in Primer ID Consensus and Raw Sequences from Nine Replicates of One Subject Sample

				Percentage and K103N	
Replicate	*Number of input viral template*	*Number of raw sequences*	*Number of consensus sequences*	*In raw sequences*	*In consensus sequences (w/95% CI)*
JKRT1	667	632,260	61	68%	69% (57–81%)
JKRT2	667	599,121	68	63%	68% (57–79%)
JKRT3	222	527,494	27	58%	56% (37–74%)
JKRT4	74	328,428	6	32%	33% (5–71%)
JKRT5	74	512,092	15	69%	60% (35–85%)
JKRT6	74	426,783	14	71%	57% (31–83%)
JKRT7	74	232,752	14	71%	71% (47–95%)
JKRT8	74	241,425	13	78%	62% (36–88%)
JKRT9	74	263,082	11	65%	64% (36–92%)
Total	2,000	3,763,437	229	64%	66% (58–70%)

Each replicate represents an independent cDNA synthesis reaction with a varying number of input HIV-1 RNA templates, estimated using sample viral load: 2 of 667 copies, 1 of 222 copies, and 6 of 74 copies. The total percentage was calculated combining the number of variants with mutations and the number of total sequences from all nine replicates.

CI, confidence interval.

platform was comparable to those obtained using the 454 FLX platform despite a nearly 10-fold increase in raw sequences. Thus, the efficiency of the deep sequencing platform was probably not a factor influencing template usage in these 17 cases. For the remaining two subjects, the number of consensus sequences increased 6- or 12-fold. Although every effort was made to sequence the same amplicons submitted for 454 sequencing, RNA extraction was repeated for these two subjects since their cDNA and amplicons were previously exhausted. In these two cases, high viral titers at or near the detection limit of the viral load assay (6.6–7.9 $\log_{10}$ HIV-1 RNA copies/ml) and dilution error could explain the discrepancy in apparent higher template utilization.

A total of 108 amino acid changes were observed in the RT coding region using both sequencing platforms, with 74 mutations detected using the 454 platform and 82 using the MiSeq platform. About 44% ($N=48/108$) were detected by both sequencing platforms, 24% ($N=26/108$) were detected by the 454 platform alone, and 31% ($N=34/108$) were detected solely by the Illumina MiSeq platform. Nearly 72% ($N=43/60$) of mutations detected by a single sequencing platform occurred on a single Primer ID consensus sequence, suggesting that these mutations were either the result of method error or of stochastic sampling of rare variants. All 34 variants detected solely by the Illumina MiSeq platform were within the downstream paired end sequence, which was revealed as an error hotspot by the control experiments (Fig. 2C). Conversely, 38% ($N=10/26$) of variants detected by the 454 platform alone were associated with homopolymeric regions, but most of the variants outside of homopolymeric regions occurred once ($N=10/16$, 62%) and could be due to a low number of templates.

To demonstrate reproducibility of the Primer ID method, multiple copy number dilutions of HIV-1 RNA from a single subject sample were tagged with unique Primer ID, amplified over HIV-1 RT codons 34–139, and submitted for Illumina MiSeq sequencing individually. We intentionally made replicates with low template input to explore the limits of the Primer ID approach in detecting minor variants. This subject had evidence of K103N on 24 of 39 (61%) Primer ID consensus sequences (454 FLX). The relative abundance of K103N was consistent across platforms, most dilutions, and resampled raw sequences versus Primer ID consensus sequences, ranging from 56% to 78% abundance (Table 2). However, the replicate JKRT4 had the lowest yield of Primer ID consensus sequences and the apparent K103N abundance was approximately half of the other replicates. Each of these replicates was represented by hundreds of thousands of raw sequence reads. Primer ID reveals the depth of sequencing, allowing an assessment of the quality of sampling of the original templates and informing the accuracy of the inference of relative abundance.

Discussion

Deep sequencing methods are subject to bias introduced by PCR amplification, and those methods that allow consecutive nucleotide additions in a homopolymeric run (e.g., 454 and Ion Torrent) are also vulnerable to erroneous calls in or near these runs.[30] Here, an alternative deep sequencing method that tags a single viral template with a unique Primer ID prior to PCR[25] was used to estimate the prevalence of preexisting RTI resistance mutations within a clinical population initiating care for HIV-1. Among 184 subjects, up to 14% had evidence of RTI resistance mutations, compared to 2.7% detection by sequencing of a bulk PCR product. An even more conservative use of the deep sequencing data based on making calls only if the mutation was associated with at least two consensus sequences gave an RTI resistance mutation detection rate of 6.0%, still more than a 2-fold increase over that seen by sequencing of a bulk PCR product, and these estimates did not include an analysis of homopolymeric regions that are susceptible to especially high error rates using the 454 sequencing platform.

Prevalence estimates must be critically interpreted since the value can be inflated due to several intrinsic errors in the sequencing methodology, not all of which can be corrected by Primer ID. The 454 platform control experiments demonstrated nearly 4-fold higher error rates within homopolymeric regions compared to heteropolymeric regions despite the use of three or more raw sequences with the same Primer ID to create a consensus sequence. Intractable homopolymeric errors argue against using sequencing platforms that are subject to these errors to estimate the prevalence of minority variants, especially those associated with homopolymeric regions such as variants with K103N and K65R.[15,17,32]

The Illumina MiSeq platform, which does not rely on the incorporation of multiple nucleotides at a homopolymeric stretch, eliminated homopolymer-associated errors in control experiments. However, this system has its own set of limitations including the accumulation of errors over the sequencing run,[33] which was consistent with our own control experiments, poor discrimination of highly similar sequences,[34] and predominance of errors in one strand of the paired ends,[34] also consistent with our results but poorly understood. When Illumina was used to sequence samples from a subset of subjects and compared to the 454 sequencing platform, concordance between the two platforms was only 50% across all queried positions, suggesting either substantial sequencing error or stochastic sampling, particularly associated with very low-abundance variants. There is some evidence of a homopolymer-associated sequencing error, since nearly 40% of mutations detected only by the 454 platform were near or within these sites. It is also likely that many of the mutations detected by the MiSeq platform alone were the result of sequencing error, since all of these mutations were within the downstream paired end, which is associated with a 3-fold increase in error. However, it is not possible to rule out stochastic sampling of the viral population as the source of the discrepancy given the limited template usage revealed by the use of Primer ID.

Our analysis of clinical subject samples was clearly limited by the number of templates we sampled, and if sufficient numbers of templates had been available (i.e., enough to give 1,000 or more consensus sequences) we could have queried down to the 0.1–0.5% range, below which residual method error still confounds the analysis. Although limited template utilization was a problem in our analysis of these samples, it was the use of Primer ID that revealed the extent of template utilization and allowed us to estimate the quality of sampling. Alternatively, if we had relied on the raw reads with an arbitrary cut-off (1%), we would have not only overestimated the prevalence of RT inhibitor resistance mutations, but we would have also erroneously concluded that our sampling depth was much higher for these samples given the number of raw reads that passed quality filters (median >2,000), and our estimates of the frequency of resistance mutations in the viral population would have been skewed upward by nearly 20% compared to Primer ID. When we repeatedly sequenced the same subject sample with a predominant homopolymer-associated K103N mutation, we observed close agreement in relative abundance between Primer ID and raw consensus sequences. However, Primer ID revealed the quality of sampling that would have been masked by relying on resampled raw sequences alone.

Even correction with Primer ID, including all resistance mutations in estimates, i.e., even those that appear in only one Primer ID consensus sequence, may fail to correct for errors introduced during cDNA synthesis, which occur in the earliest cycles of PCR amplification or which are homopolymer associated. Unfortunately, downstream data filtering with Primer ID cannot account for the first two of these biases, but control experiments using DNA as the starting template did demonstrate a substantial reduction in errors within homopolymeric regions.

In a separate deep sequencing study of homogeneous HIV sequences in which virion RNA was used as the starting template, cDNA synthesis introduced approximately one error for every 10,000 bases sequenced (S. Zhou and R. Swanstrom, unpublished observations). Sequencing errors within the Primer ID itself cannot be ruled out either, and these errors may be even more likely if the Primer ID itself contains a homopolymeric sequence. In the worst case scenario, a viral genome is linked to a homopolymeric Primer ID and subsequently oversampled, such as might occur when the number of input templates is low, and thus the number of reads of each Primer ID is high. Because the original Primer ID itself contains a homopolymeric sequence, it is more likely to be misread repeatedly and in the same way by the 454 sequencing platform. In this manner, more than one Primer ID may be linked to the same viral genome, and these would be counted as separate viral genomes when collapsed into separate consensus sequences. Most such Primer IDs are unlikely to be abundant enough to be included in consensus sequence assembly.

We assessed this type of error by building a tree of the Primer ID sequences themselves. We found no evidence of this type of oversampling in this dataset, although this type of monitoring is likely to be an important feature of using Primer ID. Primer ID may also fail when there are ties in nucleotide calls at a given position (ambiguity) among resampled raw sequences, thus making it impossible to infer the ''real'' viral sequence. This scenario is more likely to occur when consensus sequences are constructed from a low number of resampled raw sequences. Here, the number or raw reads used to generate each consensus sequence was higher than expected. For example, given 10^6 reads/454 FLX plate, 500 input cDNA for each of two amplicons, and 184 subjects, we would expect each Primer ID consensus sequence to be created from five or six resampled raw sequences on average. Over both amplicons, a median of 12 resampled raw sequences (IQR: 6–24) was used to generate each Primer ID consensus sequence. However, even with higher than expected output, we cannot exclude the contribution of platform efficiency to our ability to sample rare variants.

Despite its limitations, Primer ID offers an opportunity to make inferences about changes within the viral population when multiple resistance mutations are present, assuming that each unique Primer ID represents an individual viral genome. Only three subjects had evidence of multiple resistance mutations on non-homopolymer-associated codons, one of whom was identified by bulk sequence analysis. Two subjects had multiple, low-frequency NRTI resistance mutations on separate Primer ID consensus sequences, which could indicate past NRTI exposure followed by the reappearance of wild type from the subjects' reservoirs. The remaining subject had multiple, linked resistance mutations (M41L + L74V + K101E and Y181C + G190S + T215Y/D/S) that predominated in the Primer ID consensus sequences, with evidence of reversion at codons 74 and 215. Previous studies have shown that K101E + G190S reduce fitness compared to wild-type virus in the absence of antiretroviral therapy, but that the addition of M41L + T215Y or L74V in particular improves fitness without reducing NNRTI resistance.[35,36] In this subject it appears that T215Y is reverting more rapidly than L74V given the higher frequency of sequences with T215 revertant mutations compared to L74 (21% vs. 6%), but we are limited in our conclusions since these regions of RT were independently amplified and sequenced and thus for these different amplicons cannot be linked.

Many studies, including a recent systematic review, have linked minority pretherapy NNRTI resistance mutations with an increased risk of virologic failure.[37] Despite this evidence, questions still remain surrounding the clinical importance of minority drug-resistant variants, particularly with respect to defining a specific abundance threshold at which resistance mutations begin to affect the response to combination therapy.[13] Before any particular cut-off for clinical significance can be determined, the drug-resistant viral population must first be measured as accurately as possible. The most promising method with potential to move beyond the research setting, ultradeep sequencing, still suffers from multiple sources of error that are inherent in this method. In this study, these errors and PCR resampling were addressed using the Primer ID, which showed a 30-fold reduction in error rates over raw sequence analyses and which, despite limited viral template usage in clinical samples, still revealed additional subjects with pretherapy resistance mutations. As important, the use of Primer ID reveals the number of templates that were actually sampled, thus providing an accurate assessment of the quality of the sampling depth, an essential piece of information when evaluating the meaning of the detection of rare variants.

Acknowledgments

Funding was provided in part by the National Center for Research Resources and the National Center for Advancing Translational Sciences, National Institutes of Health, through grant UL1TR000083. The content is solely the responsibility of the authors and does not necessarily represent the official views of the NIH. Additional funding was provided by the UNC Center for AIDS Research, an NIH-funded program [P30 AI50410], and NIH award R37 AI44667. The University of North Carolina Microbiome Core Facility provided amplicon pyrosequencing with the 454 FLX platform, and the University of North Carolina High Throughput Sequencing Facility provided amplicon sequencing using the Illumina MiSeq platform.

Author Disclosure Statement

UNC is pursuing IP protection for Primer ID, and R.S. and C.J. are listed as co-inventors and have received nominal royalties.

References

1. Buchacz K, Baker RK, Palella FJ Jr, *et al.:* AIDS-defining opportunistic illnesses in US patients, 1994–2007: A cohort study. AIDS 2010;24(10):1549–1559.
2. Egger M, May M, Chene G, *et al.:* Prognosis of HIV-1-infected patients starting highly active antiretroviral therapy: A collaborative analysis of prospective studies. Lancet 2002;360(9327):119–129.
3. Moore RD, Keruly JC, and Bartlett JG: Improvement in the health of HIV-infected persons in care: Reducing disparities. Clin Infect Dis 2012;55(9):1242–1251.
4. Clavel F and Hance AJ: HIV drug resistance. N Engl J Med 2004;350(10):1023–1035.
5. Hirsch MS, Gunthard HF, Schapiro JM, *et al.:* Antiretroviral drug resistance testing in adult HIV-1 infection: 2008 Recommendations of an International AIDS Society-USA Panel. Clin Infect Dis 2008;47(2):266–285.
6. Brenner BG, Roger M, Routy JP, *et al.:* High rates of forward transmission events after acute/early HIV-1 infection. J Infect Dis 2007;195(7):951–959.
7. Yerly S, Junier T, Gayet-Ageron A, *et al.:* The impact of transmission clusters on primary drug resistance in newly diagnosed HIV-1 infection. AIDS 2009;23(11):1415–1423.
8. Bansi L, Geretti AM, Dunn D, *et al.:* Impact of transmitted drug-resistance on treatment selection and outcome of first-line highly active antiretroviral therapy (HAART). J Acquir Immune Defic Syndr 2010;53(5):633–639.
9. Poggensee G, Kucherer C, Werning J, *et al.:* Impact of transmission of drug-resistant HIV on the course of infection and the treatment success. Data from the German HIV-1 Seroconverter Study. HIV Med 2007;8(8):511–519.
10. Wittkop L, Gunthard HF, de Wolf F, *et al.:* Effect of transmitted drug resistance on virological and immunological response to initial combination antiretroviral therapy for HIV (EuroCoord-CHAIN joint project): A European multicohort study. Lancet Infect Dis 2011;11(5):363–371.
11. Panel on Antiretroviral Guidelines for Adults and Adolescents: Guidelines for the Use of Antiretroviral Agents in HIV-1-Infected Adults and Adolescents. U.S. Department of Health and Human Services. March 12, 2013. 2013:1–267. Available at www.aidsinfo.nih.gov/ContentFiles/AdultandAdolescentsGL.pdf. Accessed March 2013.
12. Palmer S, Kearney M, Maldarelli F, *et al.:* Multiple, linked human immunodeficiency virus type 1 drug resistance mutations in treatment-experienced patients are missed by standard genotype analysis. J Clin Microbiol 2005;43(1): 406–413.
13. Goodman DD, Zhou Y, Margot NA, *et al.:* Low level of the K103N HIV-1 above a threshold is associated with virological failure in treatment-naive individuals undergoing efavirenz-containing therapy. AIDS 2011;25(3): 325–333.
14. Jakobsen MR, Tolstrup M, Sogaard OS, *et al.:* Transmission of HIV-1 drug-resistant variants: Prevalence and effect on treatment outcome. Clin Infect Dis 2010;50(4): 566–573.
15. Johnson JA, Li JF, Wei X, *et al.:* Baseline detection of low-frequency drug resistance-associated mutations is strongly associated with virological failure in previously antiretroviral-naive HIV-1-infected persons. Antivir Ther 2006; 11(5):S79.
16. Lataillade M, Chiarella J, Yang R, *et al.:* Prevalence and clinical significance of HIV drug resistance mutations by ultra-deep sequencing in antiretroviral-naive subjects in the CASTLE study. PLoS One 2010;5(6):e10952.
17. Simen BB, Simons JF, Hullsiek KH, *et al.:* Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis 2009; 199(5):693–701.
18. Metzner KJ, Rauch P, Braun P, *et al.:* Prevalence of key resistance mutations K65R, K103N, and M184V as

minority HIV-1 variants in chronically HIV-1 infected, treatment-naive patients. J Clin Virol 2011;50(2):156–161.

19. Metzner KJ, Rauch P, Walter H, *et al.:* Detection of minor populations of drug-resistant HIV-1 in acute seroconverters. AIDS 2005;19(16):1819–1825.
20. Hoffmann C, Minkah N, Leipzig J, *et al.:* DNA bar coding and pyrosequencing to identify rare HIV drug resistance mutations. Nucleic Acids Res 2007;35(13):e91.
21. Gilles A, Meglecz E, Pech N, *et al.:* Accuracy and quality assessment of 454 GS-FLX titanium pyrosequencing. BMC Genom 2011;12:245.
22. Wang C, Mitsuya Y, Gharizadeh B, *et al.:* Characterization of mutation spectra with ultra-deep pyrosequencing: Application to HIV-1 drug resistance. Genome Res 2007; 17(8):1195–1201.
23. Kanagawa T: Bias and artifacts in multitemplate polymerase chain reactions (PCR). J Biosci Bioeng 2003;96(4): 317–323.
24. Liu SL, Rodrigo AG, Shankarappa R, *et al.:* HIV quasispecies and resampling. Science 1996;273(5274):415–416.
25. Jabara CB, Jones CD, Roach J, *et al.:* Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. Proc Natl Acad Sci USA 2011;108(50):20166–20171.
26. Napravnik S, Eron JJ Jr, McKaig RG, *et al.:* Factors associated with fewer visits for HIV primary care at a tertiary care center in the Southeastern U.S. AIDS Care 2006; 18(Suppl 1):S45–50.
27. Chenna R, Sugawara H, Koike T, *et al.:* Multiple sequence alignment with the Clustal series of programs. Nucleic Acids Res 2003;31(13):3497–3500.
28. Tamura K, Peterson D, Peterson N, *et al.:* MEGA5: Molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol 2011;28(10):2731–2739.
29. Bennett DE, Camacho RJ, Otelea D, *et al.:* Drug resistance mutations for surveillance of transmitted HIV-1 drug-resistance: 2009 update. PLoS One 2009;4(3):e4724.
30. Varghese V, Wang E, Babrzadeh F, *et al.:* Nucleic acid template and the risk of a PCR-induced HIV-1 drug resistance mutation. PLoS One 2010;5(6):e10992.
31. Williams RL: A note on robust variance estimation for cluster-correlated data. Biometrics 2000;56(2):645–646.
32. Kuritzkes DR, Lalama CM, Ribaudo HJ, *et al.:* Preexisting resistance to nonnucleoside reverse-transcriptase inhibitors predicts virologic failure of an efavirenz-based regimen in treatment-naive HIV-1-infected subjects. J Infect Dis 2008;197(6):867–870.
33. Bentley DR, Balasubramanian S, Swerdlow HP, *et al.:* Accurate whole human genome sequencing using reversible terminator chemistry. Nature 2008;456(7218):53–59.
34. Krueger F, Andrews SR, and Osborne CS: Large scale loss of data in low-diversity illumina sequencing libraries can be recovered by deferred cluster calling. PLoS One 2011;6(1):e16607.
35. Wang J, Bambara RA, Demeter LM, and Dykes C: Reduced fitness in cell culture of HIV-1 with nonnucleoside reverse transcriptase inhibitor-resistant mutations correlates with relative levels of reverse transcriptase content and RNase H activity in virions. J Virol 2010;84(18):9377–9389.
36. Wang J, Li D, Bambara RA, *et al.:* L74V increases the reverse transcriptase content of HIV-1 virions with non-nucleoside reverse transcriptase drug-resistant mutations L100I + K103N and K101E + G190S, which results in increased fitness. J Gen Virol 2013;94(Pt 7):1597–1607.
37. Li JZ, Paredes R, Ribaudo HJ, *et al.:* Low-frequency HIV-1 drug resistance mutations and risk of NNRTI-based antiretroviral treatment failure: A systematic review and pooled analysis. JAMA 2011;305(13):1327–1335.

Address correspondence to:
Ronald Swanstrom
22-006 Lineberger Cancer Center, Campus Box 7295
University of North Carolina at Chapel Hill
Chapel Hill, North Carolina 27599-7295

E-mail: risunc@email.unc.edu