



OPEN

A single variant sequencing method for sensitive and quantitative detection of HIV-1 minority variants

Gurjit Sidhu^{1,6}, Layla Schuster^{1,2,6}, Lin Liu^{1,5}, Ryan Tamashiro¹, Eric Li¹, Taimour Langae³, Richard Wagner² & Gary P. Wang^{1,4} ✉

HIV drug resistance is a major threat to achieving long-term viral suppression in HIV-positive individuals. Drug resistant HIV variants, including minority variants, can compromise response to antiretroviral therapy. Many studies have investigated the clinical relevance of drug resistant minority variants, but the level at which minority variants become clinically relevant remains unclear. A combination of Primer-ID and deep sequencing is a promising approach that may quantify minority variants more accurately compared to standard deep sequencing. However, most studies that used the Primer-ID method have analyzed clinical samples directly. Thus, its sensitivity and quantitative accuracy have not been adequately validated using known controls. Here, we constructed defined proportions of artificial RNA and virus quasispecies and measured their relative proportions using the Primer-ID based, quantitative single-variant sequencing (qSVS) assay. Our results showed that minority variants present at 1% of quasispecies were detected reproducibly with minimal variations between technical replicates. In addition, the measured frequencies were comparable to the expected frequencies. These data validate the accuracy and reproducibility of the qSVS assay in quantifying authentic HIV minority variants, and support the use of this approach to examine the impacts of minority HIV variants on virologic response and clinical outcome.

There are approximately 37.9 million people living with HIV^{1,2}. As antiretroviral drug coverage increases and patients become more treatment experienced, the prevalence of HIV drug resistance is expected to increase³. The presence of drug resistant HIV in the viral population is known to compromise virologic response to ART⁴⁻⁶. Currently, genotypic and phenotypic assays are used to identify HIV drug resistance in clinical settings. However, because phenotypic assays are time-consuming and expensive, they are generally reserved for patients with known or suspected complex drug-resistance patterns. In contrast, genotypic testing by Sanger Sequencing is the preferred approach for most patients beginning or switching ART. While the Sanger Sequencing method is informative in many settings, it is insensitive in detecting minority variants circulating in less than 20% of the viral quasispecies^{7,8}. Many studies have shown that minority drug resistant variants can compromise response to therapy^{5,6,9-22}. However, the level at which minority variants become clinically relevant remains unclear.

Numerous studies have used deep sequencing methods such as Roche/454 pyrosequencing and Illumina sequencing to detect minority HIV variants²³⁻³⁰. However, even with the next generation sequencing (NGS) technology, the original quasispecies structure may be distorted by technical artifacts during sample preparation, which may incorporate errors during reverse transcription and PCR, resampling errors from low copy number of viral templates, PCR amplification bias, and sequencing errors from Roche/454 or Illumina sequencing. Thus, quantification of minority variants using standard deep sequencing approaches may not be accurate.

Several years ago, Jabara *et al.*³¹ developed a Primer-ID approach which solved many of the technical artifacts and biases described above. This approach relies on the use of random sequence tags in the reverse transcription primer such that each RNA template from the original quasispecies population receives a unique Primer-ID. After sequencing, datasets are clustered according to the unique Primer-ID, and a consensus sequence for each

¹Division of Infectious Disease and Global Medicine, Department of Medicine, University of Florida, Gainesville, FL, USA. ²Medosome Biosciences, Alachua, FL, USA. ³Department of Pharmacotherapy and Translational Research, Center for Pharmacogenomics and Precision Medicine, College of Pharmacy, University of Florida, Gainesville, FL, USA. ⁴Infectious Diseases Section, Medical Service, North Florida/South Georgia Veterans Health System, Gainesville, FL, USA. ⁵Present address: Department of Medicine, St. Luke's Hospital, Chesterfield, MO, USA. ⁶These authors contributed equally: Gurjit Sidhu and Layla Shuster. ✉e-mail: gary.wang@medicine.ufl.edu

template is generated. Since each starting RNA template is tagged with a unique Primer-ID barcode, errors originating from subsequent steps such as nucleotide misincorporation and deep sequencing, allelic skewing from differential amplification, and template resampling, are corrected using this approach. Thus, the resulting population of consensus sequences should accurately represent the original quasispecies population, and the approach holds great promise for quantifying minority variants in clinical samples. However, most published studies have applied the Primer-ID method directly to clinical samples^{27,32–37}. The sensitivity, reproducibility, and quantitative accuracy of the approach have not been adequately validated using quasispecies of known proportions.

In the present study, we constructed artificial RNA pools and artificial viral quasispecies and determined the frequency of minority HIV variants using a Primer-ID, quantitative single variant sequencing assay. The goal was to validate using known controls that the assay can detect minority variants reproducibly at or above 1% of the HIV-1 quasispecies with high precision.

Results

Determination of the background error rate. The single variant sequencing (SVS) method leveraged the Primer-ID approach and Illumina sequencing to quantify the abundance of different variants in viral quasispecies. We first determined the background error rate for the SVS method by selecting three plasmids containing the entire HIV-1 PR gene and three plasmids containing partial RT gene segment (codon 1–237) and subjecting each individual plasmid to *in vitro* transcription, PCR amplification, and deep sequencing using the SVS procedure (Fig. 2). We then selected three cell culture-derived HIV and subjected each virus stock to the same SVS analysis. SVS analysis was performed for the PR and RT genes for plasmids and the PR gene for viruses. Amino acids called erroneously were identified and their frequency determined at each position (see Supplemental Tables ST1 and ST2). Mean of error frequencies at each position of the given sequence was calculated and reported as the background error rate per position for that particular gene. For both plasmid-derived RNA and viral RNA, the background error rate per position was found to be less than 0.1% in all samples (Fig. 2).

Sensitive and quantitative detection of individual variants in plasmid-derived RNA quasispecies. To determine the threshold at which individual minority variants could be detected consistently, we generated artificial pools of RNA quasispecies derived from plasmids of known sequences. The concentration of each linearized plasmid was determined, and combined in different ratios to generate 3 plasmid populations with defined proportions (Supplemental Table ST4). Two of the three populations included minority species at 1% mean abundance (i.e. pools B and C for PR; pools E and F for RT). RNA quasispecies was then generated by *in vitro* transcription of the plasmid pools. To evaluate the contribution of pipetting variations during the construction of artificial pools, each pool was prepared in quadruplicates by manual pipetting and also by a liquid handling robot.

The abundance of individual variants in each PR or RT artificial RNA pool is shown in Fig. 3. The expected (theoretical) frequency of individual variants in each pool was calculated based on measured concentrations of linearized plasmids. The SVS analysis revealed that minority variants present at mean 1% abundance of the quasispecies populations (i.e. Protease pool B p50V, Protease pool C p84V, Reverse Transcriptase pool E p82A, and Reverse Transcriptase Pool C p84V) were detected in all replicates, although the measured mean abundance deviated from the expected abundance in some cases (observed mean abundance - Protease pool B p50V: 0.14%, Protease pool C p84V: 1.70%, Reverse Transcriptase pool E p82A: 0.08%, and Reverse Transcriptase Pool F p151M: 0.33%). We observed minimal differences between technical replicates, and also between the two pipetting methods (liquid handling robot vs manual pipetting). These results indicate that the SVS method reproducibly detects minority variants present at or above 1% abundance of quasispecies.

Sensitive and quantitative detection of individual variants in virus-derived RNA quasispecies. To demonstrate that the SVS method can be applied to viral populations, we selected three cell culture-derived viruses that differed in two positions (82 and 84) in the PR gene and constructed four artificial viral mixtures with defined proportions (Supplemental Table ST5). Each artificial pool consists of two minority variants ranging from 0.8–12.7% and one majority variant (>84.1%). The abundance of each virus was determined using SVS and compared with the expected (theoretical) abundance calculated based on p24 titers. The analysis showed that minority variants (i.e. control virus) present at 1 to 3% of the expected frequency were detected reproducibly and were comparable to the expected frequencies (4.49% vs 3.21% expected, 3.23% vs 2.38% expected, 2.41% vs 1.57% expected, 0.84% vs 0.77% expected; mix1 to mix4, respectively) (Fig. 4). In addition, variations in measured frequencies between technical replicates were small. These results demonstrate that the SVS procedure reproducibly detected minority variants at or above ~1% of the viral quasispecies.

To further confirm the high sensitivity and accuracy of the SVS assay, the frequencies of authentic minority variants in virus Mix 4 and Mix 3 were compared to the frequencies of reads that were called erroneously (Fig. 5). The frequency of authentic minority variants at position 82 and 84 called correctly (green bars; 0.77% and 2.41% in virus Mix 4 and Mix 3, respectively) was significantly higher than the frequencies of erroneous calls (variant with the highest frequency has a mean abundance of 0.143% for Mix 4 and 0.158% for Mix 3; $p < 0.00001$). Overall, >99.6% of the reads were called correctly in all four mixes. These results indicate that authentic minority variants present in as low as 0.8% frequency in the viral quasispecies could be distinguished accurately from the background errors.

Discussion

HIV-1 drug resistant minority variants can compromise response to antiretroviral therapy. Although many studies have investigated the clinical impact of minority variants, the clinical significance of HIV-1 drug resistant minority variants continues to be debated^{27,29,30,37–40}. Furthermore, the level at which drug resistant minority variants become clinically relevant remains unclear. The Primer-ID method is one approach that can reduce

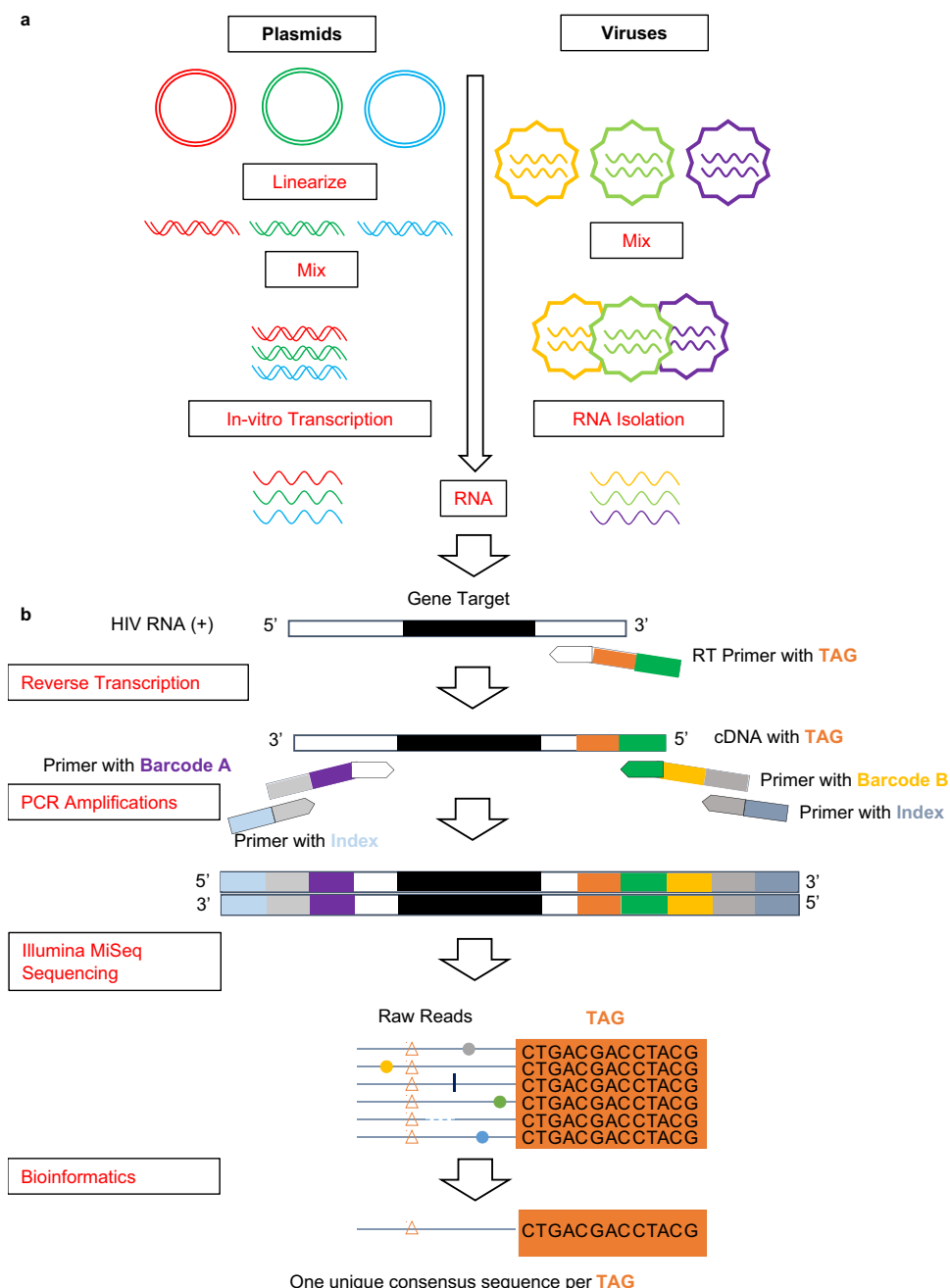


Figure 1. (a) Generation of artificial pools of RNA quasispecies or virus quasispecies. (left) Plasmids encoding PR or RT gene segments were linearized, then mixed in different proportions to generate artificial pools. RNA quasispecies was generated by *in vitro* transcription. (right) Viral particles were mixed in different proportions to generate virus quasispecies, then viral RNA was extracted. (b) The frequency of each RNA quasispecies was quantified using single variant sequencing, leveraging the primer-ID approach and high-throughput Illumina sequencing.

accumulation of erroneous variants⁴¹. Correctly calling minority variants while simultaneously reducing erroneous base calls is paramount for accurate determination of minority drug-resistance mutations and studies of their impacts on treatment response. Here, we leveraged the Primer-ID approach and Illumina deep sequencing and report the validation of a quantitative Single Variant Sequencing (SVS) method for accurate and sensitive detection of minority HIV-1 variants. We constructed artificial RNA quasispecies of defined proportions by *in vitro* transcription of plasmids and direct extraction of cell culture-derived viruses, then interrogated the proportions of RNA quasispecies using the SVS methods.

To accurately identify and quantify authentic HIV-1 minority variants, it is critical that an assay can reliably distinguish authentic low frequency variants from errors generated during nucleic acid amplification and the deep sequencing process. We first mixed plasmids carrying HIV-1 PR and RT gene segments with known

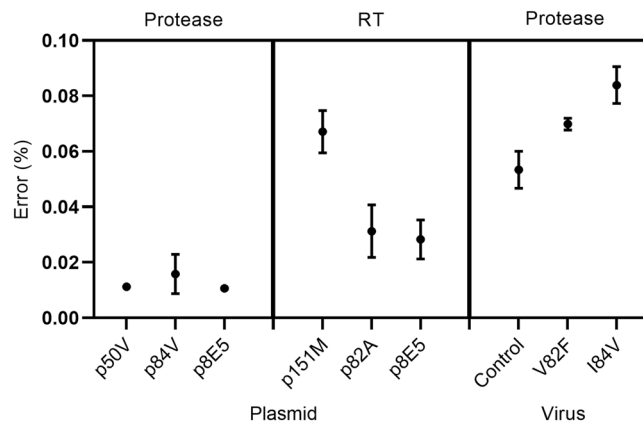


Figure 2. Background error rate of the single variant sequencing (SVS) method. Purified plasmids (in triplicate) or viruses (in duplicate) were subjected to the SVS analysis. For each of the three plasmids, RNA encoding HIV-1 PR (amino acids 8–99) or RT (amino acids 11–133) was generated by *in vitro* transcription, followed by the SVS procedure and Illumina sequencing. For each of the three virus stocks, viral RNA was extracted, then subjected to SVS and Illumina sequencing of the PR gene segment (amino acids 8–99). Frequency of amino acids called erroneously at all sequenced positions was scored and plotted as % mean error per position for the sequenced gene. Mean with SEM is shown.

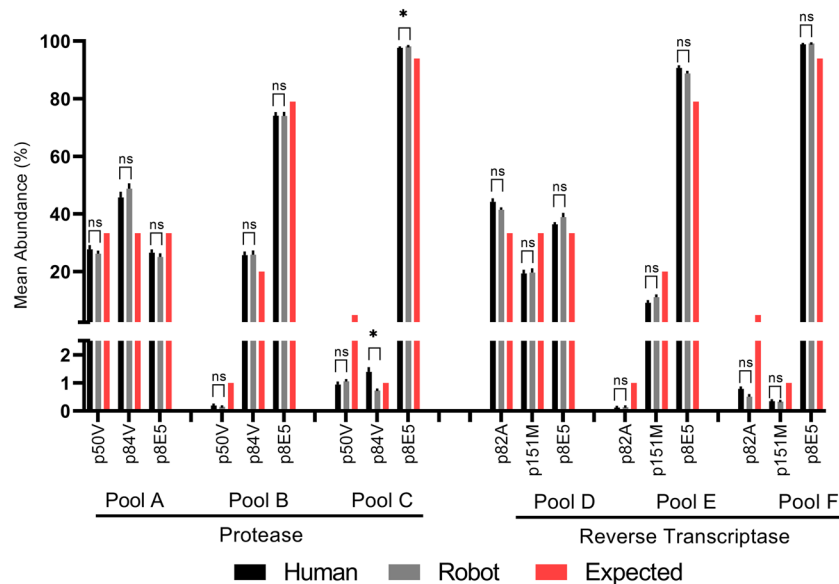


Figure 3. Frequency of individual variants in quasispecies pools determined using the SVS procedure. Each artificial RNA quasispecies is shown as a panel (pools A–F). Each bar within the panel represents the abundance of an individual variant, and the red bar denotes the expected frequency calculated based on the initial plasmid concentration. Each plasmid pool was prepared in quadruplicates using a liquid handling robot (gray) or manual pipetting (black). Individual variants in the PR (left) or RT (right) genes were analyzed and the mean abundance (%) with SEM of each variant was plotted. ns = not statistically significant. * = $p < 0.005$.

mutations, then generated RNA quasispecies by *in vitro* transcription of the plasmid pools. Expected proportions of different plasmids were calculated based on plasmid concentrations to include one or two minority variants (less than 20%) and one or two majority variants (20% or higher). Our results showed that the assay was highly sensitive in accurately detecting minority variants in as low as 1% of the quasispecies population, with a low background error rate of $<0.1\%$. The low error rate was consistent with a recent report by Howison *et al.* that also used the Primer ID method⁴¹. Importantly, our data demonstrated a high level of reproducibility among technical replicates.

In some cases, the abundance of variants (both majority and minority variants) deviated from the expected abundance calculated based on measured plasmid concentrations. The deviations for individual quasispecies were consistent among technical replicates in all quasispecies (i.e. reproducibly higher or lower than the expected values and with a similar magnitude) (Fig. 3). We speculate that the observed differences were a result

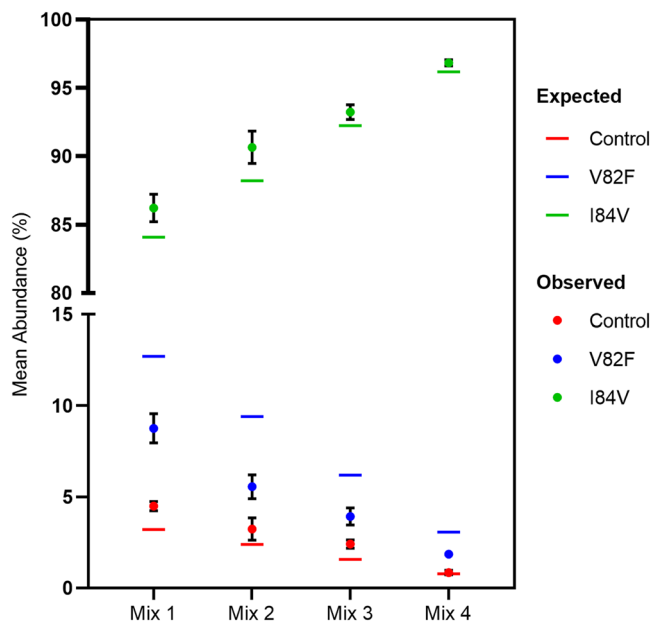


Figure 4. Frequency of individual variants in virus quasispecies determined using SVS. Mean abundance (%) of each variant in the artificial viral quasispecies (mix 1–4) is shown on the y-axis. The horizontal line denotes the expected frequency. Each virus pool was prepared in quadruplicates and amplified and sequenced independently. Mean with SEM is shown.

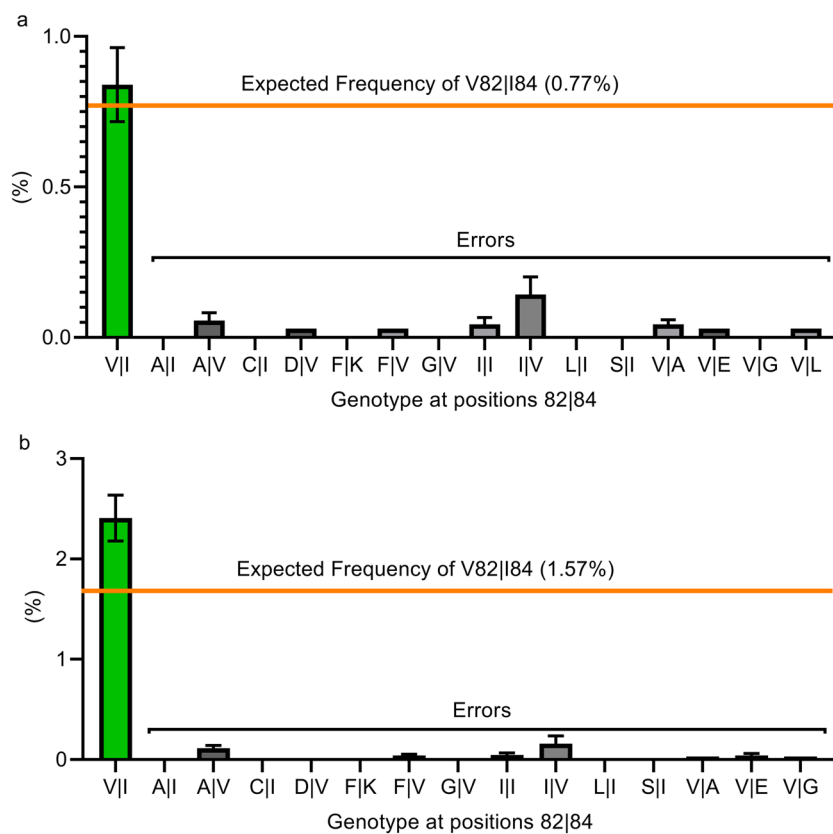


Figure 5. Detection of low frequency authentic minority variants in viral quasispecies. **(a)** Observed mean abundance (%) (green bar) vs. expected frequency (orange line) of Control (V82|I84) virus in Mix 4. **(b)** Observed mean abundance (%) (green bar) vs. expected frequency (orange line) of Control (V82|I84) virus in Mix 3. Gray bars represent the frequency of viruses with amino acids called incorrectly at positions 82 and 84. Each sample was performed in quadruplicate. Mean with SEM was plotted.

of concentration measurements of linearized plasmids, and/or stochastic events from *in-vitro* transcription (i.e. differences in the amount of RNA transcribed *in vitro* from different plasmids). Seifert *et al.*⁴² combined Primer ID with MiSeq sequencing to study heterogeneous HIV-1 populations by mixing five viruses in equal proportions (20% each) calculated based on RNA copy numbers. However, they observed proportions ranging between 6% and 38%. Similar deviations were reported by Howison *et al.*⁴¹ in their analysis of artificial quasispecies based on RT-PCR or plasmid DNA measurements. Taken together, these data suggest that *in vitro* plasmid or RNA-based quasispecies may be adequate for the evaluation of assay sensitivity, but may not be sufficiently robust for determining the quantitative accuracy of variant populations in quasispecies. Going forward, an improved method for constructing accurate proportions of viral quasispecies is essential for validating the quantitative accuracy of the SVS assay.

To demonstrate that the SVS method is also applicable to viral populations, we generated defined mixtures of HIV-1 virus quasispecies. To minimize the impact of concentration measurements on accurate proportions of each viral population, we first combined two variants at 1:4 ratio, and then spiked the mixture into the third virus variant to generate four viral quasispecies consisting of two minority variants in a fixed ratio (1:4) ranging from 0.8% to 12.7% abundance and one majority variant at >84% abundance. This approach was expected to generate different quasispecies with a fixed ratio between the two minority variants. Indeed, our data (Fig. 4) demonstrated a consistent ratio between the two minority variants across all four viral quasispecies, with the observed ratio (1:2) deviating slightly from the expected value (1:4). This high level of consistency across the four quasispecies suggests that the differences likely resulted from an over or under estimation of viral RNA copy number based on p24 measurements. Overall, the SVS assay demonstrated high reproducibility in identifying and quantifying both majority and minority variants across different sample types with minor errors that do not impact data interpretation. Importantly, these results demonstrate that the SVS method can consistently quantify authentic minority variants present in as low as 0.8% of the viral quasispecies.

This study has several limitations. First, the initial Primer-IDs label individual RNA templates during reverse transcription. The SuperScript IV Reverse Transcriptase possesses no proofreading activity and may have introduced errors during the first strand cDNA synthesis which could not be corrected in subsequent steps⁴³. Second, PCR amplification following cDNA synthesis may have introduced additional errors in the Primer-ID barcode sequence, thereby generating offspring Primer-IDs, the effect of which could be minimized using a cutoff model proposed by Zhou *et al.*⁴⁴. However, the use of high-fidelity Platinum SuperFi DNA Polymerase should reduce the likelihood of additional errors introduced during the PCR step.

In summary, we have validated the SVS method using known controls, and showed that the assay consistently detects minority variants at or above 1% level with high precision. These results support the use of quantitative single variant sequencing assay to examine the impacts of minority HIV-1 drug resistant variants on virologic response and clinical outcome.

Methods

Construction of artificial RNA pools with defined proportions. Plasmids (p8E5, p50V, p84V, p82A and p151M) with unique amino acid polymorphism (Supplemental Table ST7) were obtained from NIH AIDS Research and Reference Reagent Program. These constructs carry a 1060-bp fragment spanning from *gag* gene to codon 237 of the reverse transcriptase gene (nucleotide 2201 through 3261) of the HIV-1 genome (GenBank accession no. K03455) in pCR 2.1-TOPO vector (Invitrogen). To confirm the sequence of the insert, plasmids were transformed into *E. coli* (One Shot TOP10 Competent Cells, ThermoFisher, Cat # C404010), single colonies were selected and grown in Luria Broth with 25 µg/ml kanamycin, and plasmid DNA was isolated using NucleoSpin Plasmid (NoLid) kit (Macherey-Nage, Cat # 740499). The sequence of the PR and RT gene segments in each plasmid was confirmed by Sanger sequencing using following primers; PRC (forward): 5'-CTCCCCCTCAGAAGCAGGAGCCGATAGACAAGGAAGTGTATCC and RT3 (reverse): 5'-TATCAGGATGGAGTTCATAAC. Next, plasmids were linearized using *Bam*HI-HF restriction enzyme (New England Biolabs), and the length of the linearized plasmid was verified by agarose gel electrophoresis and DNA concentration measured using the Qubit dsDNA HS Assay Kit in Qubit 4 fluorometer (Invitrogen). Purified linearized plasmids were mixed to create 12 pools with defined proportions based on measured DNA concentrations (Supplemental Table ST4; Fig. 1a). Six pools were prepared by hand pipetting, and six identical pools were constructed by robot pipetting using Eppendorf epMotion 5070 liquid handling robot.

To generate artificial RNA pools with defined proportions (Supplemental Table ST4; Fig. 1a), individual plasmids or plasmid pools were transcribed *in vitro* using T7 RiboMAX Express Large Scale RNA Production System (Promega, Cat # P1320), and the transcribed RNA was purified using the NucleoSpin RNA Clean-up Kit (Macherey-Nagel, Cat # 740948). The RNA concentration was measured using the Nanodrop 2000 (Thermo Fisher) and adjusted to 100,000 copies/µL for subsequent reverse transcription and PCR.

Construction of virus pools of defined proportions. HIV-1_{RF} (Cat# 2806; Control; V82, I84), HIV-1_{RF/L323-12-3} (Cat# 2806; V82F) and HIV-1_{RF/L323-9-1} (Cat# 2807; I84V) (cell-free culture supernatant) were obtained from the NIH AIDS Reagent Program, Division of AIDS, NIAID, NIH^{45,46}, contributed by Dr. Dean Winslow. To construct the virus pools with defined proportions, the p24 content was determined using the HIV1 p24 ELISA Kit (Abcam, Cat # ab218268), then 10 µL each from Control (#2803) and V82F (#2806) were combined to obtain a mixture with 1 to 3.96 ratio (Control and V82F) based on p24, then this mixture was added in differing ratios to I84F (#2807) to generate four virus pools with known proportions (Supplemental Table ST5; Fig. 1a). Viral RNA was isolated using the QIAamp Viral RNA Mini Kit (Qiagen, Cat # 52904).

Single variant sequencing (SVS). A schematic of the SVS procedure is shown in Fig. 1b, which included reverse transcription, PCR amplification, pooling, and Illumina sequencing. The SVS procedure utilizes a unique

design of reverse transcription (RT) primers for priming the first strand of cDNA synthesis. Each RT primer molecule includes a 14 random nucleotide sequence constituting a unique Primer-ID Tag, which is flanked by a sequence at the 3' end that anneals to the RNA template and a sequence at the 5' end that serves as the annealing site for PCR primers (Supplemental Table ST6). The RNA from *in vitro* transcription or RNA isolated from virus particles (Fig. 1a) serves as the template for cDNA synthesis using SuperScript IV Reverse Transcriptase (SuperScript IV First-Strand Synthesis System, Invitrogen) and amplicon (protease or reverse transcriptase) specific RT primers that contained the Primer-ID tag. Each reaction contained 10^2 – 10^3 fold molar excess of primers to ensure that each RNA template was reverse transcribed to generate cDNA labeled with a unique primer ID tag. The resulting cDNA was purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Cat# 740609), then amplified using Platinum SuperFi DNA polymerase (Invitrogen) with forward and reverse PCR primers that included 4–8 variable length nucleotide barcode sequences specific to each sample. The PCR products were separated by agarose gel electrophoresis, purified using NucleoSpin Gel and PCR Clean-up (Macherey-Nagel, Cat# 740609), and tailed with index sequences required for Illumina sequencing. Amplified DNA were combined in equimolar pool, gel purified and quantified by qPCR (KAPA Library Quantification kit for Illumina sequencing platforms, Kapa Biosystems), then sequenced on the Illumina MiSeq using the v3 600 cycle kit and a 20% PhiX spike-in. Protease (PR) gene segment (amino acids 8–99) and reverse transcriptase (RT) gene segment (amino acids 11–133) were sequenced. SVS analysis was performed for the PR and RT genes for plasmids and the PR gene for viruses.

Bioinformatics. Raw Illumina MiSeq. 301 bp x 2 reads were de-multiplexed into individual samples according to the unique variable length barcode combination (4bp–8bp in length) on each end. Additional filtering criteria included an exact match to PCR primer sequences, an average quality score of 30 or higher (<0.001 error rate in raw reads), and a minimum length of 270 bp for each paired-end read. Each paired-end read was joined using FLASH (<http://ccb.jhu.edu/software/FLASH/>) with a minimum of 10 base overlap. For each sample, the joined reads were grouped by the unique 19-bp tags (including 5 bp control bases) introduced during reverse transcription. Consensus sequences were called for each unique tag via alignment using MAFFT (<http://mafft.cbrc.jp/alignment/software/>) when three or more reads share the same tag. Consensus sequences that contained ties at certain positions resulting in degenerate bases were excluded due to ambiguity. The resulting consensus sequences were aligned against corresponding reference sequences and manually inspected and corrected for artificial gaps introduced from the sequencing process. Translation of codons and summarization of mutations were carried out using custom scripts in R (<https://www.r-project.org/>) with the BioStrings package (<http://bioconductor.org/packages/release/bioc/html/Biostrings.html>).

Received: 3 December 2019; Accepted: 20 April 2020;

Published online: 18 May 2020

References

- UNAIDS. Communities at the centre, global AIDS update 2019. Defending rights, breaking barriers, reaching people with HIV services. *Unaid Data Updat.*, <https://doi.org/10.2307/j.ctt1t898kc.12> (2019).
- UNAIDS. *Fact sheet - Global AIDS UPDATE 2019*. *Unaid* (2019).
- Pham, Q. D., Wilson, D. P., Law, M. G., Kelleher, A. D. & Zhang, L. Global burden of transmitted HIV drug resistance and HIV-exposure categories. *Aids* **28**, 2751–2762 (2014).
- Johnson, J. A. *et al.* Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naïve populations and associate with reduced treatment efficacy. *PLoS Med.* **5**, 1112–1122 (2008).
- Simen, B. B. *et al.* Low-Abundance Drug-Resistant Viral Variants in Chronically HIV-Infected, Antiretroviral Treatment-Naïve Patients Significantly Impact Treatment Outcomes. *J. Infect. Dis.* **199**, 693–701 (2009).
- Paredes, R. *et al.* Pre-existing Minority Drug-Resistant HIV-1 Variants, Adherence, and Risk of Antiretroviral Treatment Failure. *J. Infect. Dis.* **201**, 100126095936095–000 (2010).
- Larder, B. A. *et al.* Quantitative detection of HIV-1 drug resistance mutations by automated DNA sequencing. *Nature* **365**, 671–673 (1993).
- Church, J. D. *et al.* Sensitivity of the ViroSeq HIV-1 Genotyping System for Detection of the K103N Resistance Mutation in HIV-1 Subtypes A, C, and D. *J. Mol. Diagnostics* **8**, 430–432 (2006).
- Goodman, D. D. *et al.* Low level of the K103N HIV-1 above a threshold is associated with virological failure in treatment-naïve individuals undergoing efavirenz-containing therapy. *Aids* **25**, 325–333 (2011).
- Balduin, M. *et al.* Prevalence of minor variants of HIV strains at reverse transcriptase position 103 in therapy-naïve patients and their impact on the virological failure. *J. Clin. Virol.* **45**, 34–38 (2009).
- Archer, J. *et al.* The evolutionary analysis of emerging low frequency HIV-1 CXCR4 using variants through time-an ultra-deep approach. *PLoS Comput. Biol.* **6**, e1001022 (2010).
- Tsibris, A. M. N. *et al.* Quantitative Deep Sequencing Reveals Dynamic HIV-1 Escape and Large Population Shifts during CCR5 Antagonist Therapy In Vivo. *PLoS One* **4**, e5683 (2009).
- Archer, J. *et al.* Detection of low-frequency pretherapy chemokine (CXC motif) receptor 4 (CXCR4)-using HIV-1 with ultra-deep pyrosequencing. *Aids* **23**, 1209–1218 (2009).
- Westby, M. *et al.* Emergence of CXCR4-Using Human Immunodeficiency Virus Type 1 (HIV-1) Variants in a Minority of HIV-1-Infected Patients following Treatment with the CCR5 Antagonist Maraviroc Is from a Pretreatment CXCR4-Using Virus Reservoir. *J. Virol.* **80**, 4909 LP–4920 (2006).
- Cozzi-Lepri, A. *et al.* Low-frequency drug-resistant HIV-1 and risk of virological failure to first-line NNRTI-based ART: A multicohort European case-control study using centralized ultrasensitive 454 pyrosequencing. *J. Antimicrob. Chemother.*, <https://doi.org/10.1093/jac/dku426> (2015).
- Jackson, J. B. *et al.* Identification of the K103N resistance mutation in Ugandan women receiving nevirapine to prevent HIV-1 vertical transmission. *Aids* **14**, F111–F115 (2000).
- Johnson, J. A. *et al.* Emergence of Drug-Resistant HIV-1 after Intrapartum Administration of Single-Dose Nevirapine Is Substantially Underestimated. *J. Infect. Dis.* **192**, 16–23 (2005).
- Church, J. D. *et al.* Sensitivity of the ViroSeq HIV-1 Genotyping System for detection of the K103N resistance mutation in HIV-1 subtypes A, C, and D. *J. Mol. Diagnostics* **8**, 430–432 (2006).

19. Hauser, A. *et al.* Detection and quantification of minor human immunodeficiency virus type 1 variants harboring K103N and Y181C resistance mutations in subtype A and D isolates by allele-specific real-time PCR. *Antimicrob. Agents Chemother.* **53**, 2965–2973 (2009).
20. Li, J. Z. *et al.* Low-Frequency HIV-1 Drug Resistance Mutations and Risk of NNRTI-Based Antiretroviral Treatment Failure: A Systematic Review and Pooled Analysis. *JAMA* **305**, 1327–1335 (2011).
21. Peuchant, O. *et al.* Transmission of HIV-1 minority-resistant variants and response to first-line antiretroviral therapy. *Aids* **22**, 1417–1423 (2008).
22. Metzner, K. J. *et al.* Minority Quasispecies of Drug-Resistant HIV-1 That Lead to Early Therapy Failure in Treatment-Naive and -Adherent Patients. *Clin. Infect. Dis.* **48**, 239–247 (2009).
23. Fred, K. *et al.* Low-Frequency Drug Resistance in HIV-Infected Ugandans on Antiretroviral Treatment Is Associated with Regimen Failure. *Antimicrob. Agents Chemother.* **60**, 3380–3397 (2016).
24. SahBandar, I. N. *et al.* Ultra-Deep Sequencing Analysis on HIV Drug-Resistance-Associated Mutations Among HIV-Infected Individuals: First Report from the Philippines. *AIDS Res. Hum. Retroviruses* **33**, 1099–1106 (2017).
25. Alves, B. M. *et al.* Characterization of HIV-1 near full-length proviral genome quasispecies from patients with undetectable viral load undergoing first-line HAART therapy. *Viruses* **9**, 392 (2017).
26. Moscona, R. *et al.* Comparison between next-generation and Sanger-based sequencing for the detection of transmitted drug-resistance mutations among recently infected HIV-1 patients in Israel, 2000–2014. *J. Int. AIDS Soc* **20**, 21846 (2017).
27. Raymond, S. *et al.* Impact of Human Immunodeficiency Virus Type 1 Minority Variants on the Virus Response to a Rilpivirine-Based First-line Regimen. *Clin. Infect. Dis.* **66**, 1588–1594 (2018).
28. Tzou, P. L. *et al.* Comparison of an In Vitro Diagnostic Next-Generation Sequencing Assay with Sanger Sequencing for HIV-1 Genotypic Resistance Testing. *J. Clin. Microbiol.* **56**, 1–13 (2018).
29. Silver, N. *et al.* Characterization of minority HIV-1 drug resistant variants in the United Kingdom following the verification of a deep sequencing-based HIV-1 genotyping and tropism assay. *AIDS Res. Ther* **15**, 1–18 (2018).
30. Inzaule, S. C. *et al.* Clinically relevant thresholds for ultrasensitive HIV drug resistance testing: a multi-country nested case-control study. *Lancet HIV* **5**, e638–e646 (2018).
31. Jabara, C. B., Jones, C. D., Roach, J., Anderson, J. A. & Swanstrom, R. Accurate sampling and deep sequencing of the HIV-1 protease gene using a Primer ID. *Proc. Natl. Acad. Sci.* **108**, 20166–20171 (2011).
32. Dennis, A. M. *et al.* Using Primer-ID Deep Sequencing to Detect Recent HIV-1 Infection. *J. Infect. Dis.* 1–6, <https://doi.org/10.1093/infdis/jiy426/5054087> (2018).
33. Keys, J. R. *et al.* Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain. *AIDS Res. Hum. Retroviruses* **31**, 658–668 (2015).
34. Zhou, S., Bednar, M. M., Sturdevant, C. B., Hauser, B. M. & Swanstrom, R. Deep Sequencing of the HIV-1 *env* Gene Reveals Discrete X4 Lineages and Linkage Disequilibrium between X4 and R5 Viruses in the V1/V2 and V3 Variable Regions. *J. Virol.* **90**, 7142–7158 (2016).
35. Bednar, M. M. *et al.* Diversity and Tropism of HIV-1 Rebound Virus Populations in Plasma Level After Treatment Discontinuation. *J. Infect. Dis.* **214**, 403–407 (2016).
36. Kearney, M. F. *et al.* Lower pre-ART intra-participant HIV-1 pol diversity may not be associated with virologic failure in adults. *PLoS One* **13**, 1–12 (2018).
37. Hassan, A. S. *et al.* Presence, persistence and effects of pretreatment HIV-1 drug resistance variants detected using next generation sequencing: A Retrospective longitudinal study from rural coastal Kenya. *PLoS One* **14**, 1–16 (2019).
38. Dalmat, R. R. *et al.* Limited Marginal Utility of Deep Sequencing for HIV Drug Resistance Testing in the Age of Integrase Inhibitors. *J. Clin. Microbiol.* **56**, 1–13 (2018).
39. Su, B. *et al.* Detection of pretreatment minority HIV-1 reverse transcriptase inhibitor-resistant variants by ultra-deep sequencing has a limited impact on virological outcomes. *J. Antimicrob. Chemother* **74**, 1408–1416 (2019).
40. Clutter, D. S. *et al.* Prevalence of drug-resistant minority variants in untreated HIV-1-infected individuals with and those without transmitted drug resistance detected by sanger sequencing. *J. Infect. Dis* **216**, 387–391 (2017).
41. Howison, M., Coetzer, M. & Kantor, R. Measurement error and variant-calling in deep Illumina sequencing of HIV. *Bioinformatics* **35**, 2029–2035 (2019).
42. Seifert, D. *et al.* A Comprehensive Analysis of Primer IDs to Study Heterogeneous HIV-1 Populations. *J. Mol. Biol.* **428**, 238–250 (2016).
43. Taylor, T. *et al.* A MiSeq-HyDRA platform for enhanced HIV drug resistance genotyping and surveillance. *Sci. Rep* **9**, 8970 (2019).
44. Zhou, S., Jones, C., Mieczkowski, P. & Swanstrom, R. Primer ID Validates Template Sampling Depth and Greatly Reduces the Error Rate of Next-Generation Sequencing of HIV-1 Genomic RNA Populations. *J. Virol.* **89**, 8540–8555 (2015).
45. Otto, M. J. *et al.* In vitro Isolation and Identification of Human Immunodeficiency Virus (HIV) Variants with Reduced Sensitivity to C-2 Symmetrical Inhibitors of HIV Type 1 Protease. *Proc. Natl. Acad. Sci. U. S. A.* **90**, 7543–7547 (1993).
46. King, R. W. *et al.* Multiple Mutations in the Human Immunodeficiency Virus Protease Gene Are Responsible for Decreased Susceptibility to Protease Inhibitors. *Antivir. Chem. Chemother.* **6**, 80–88 (1995).

Acknowledgements

We thank members of Wang laboratory for helpful discussion. This work was supported by NIAID R41AI122855, and in part by NIAID R43DK112540 and the Gatorade Trust through funds distributed by the University of Florida, Department of Medicine.

Author contributions

G.P.W. conceived the study. G.S., L.L., G.P.W. designed the experiments. G.S., L.S., L.L., T.L. performed the experiments. G.S., L.S., L.L., R.T., E.L., G.P.W. analyzed and interpreted the data. T.L. contributed analysis tools. G.S., G.P.W. drafted the manuscript. G.S., L.L., R.T., T.L., R.W., G.P.W. critically revised the manuscript.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-65085-y>.

Correspondence and requests for materials should be addressed to G.P.W.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020