Utility of Next-Generation Sequencing for HIV Antiretroviral Resistance Genotyping in a Pragmatic Clinical Laboratory Setting

Ronit Dalmat

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Public Health

University of Washington

2018

Committee:

Anna Wald

Joshua Herbeck

Program Authorized to Offer Degree:

Epidemiology

University of Washington

Abstract

Utility of Next-Generation Sequencing for HIV Antiretroviral Resistance Genotyping in a Pragmatic Clinical Laboratory Setting

Ronit Dalmat

Chair of the Supervisory Committee:

Anna Wald

Departments of Medicine, Epidemiology, and Laboratory Medicine

Routine HIV antiviral resistance genotyping is an important tool in the clinical management of HIV infections and interest is growing among clinical laboratories for the use of next-generation sequencing (NGS) in HIV drug resistance testing. Although resistance genotyping has traditionally been conducted using Sanger sequencing methods, NGS is emerging as a powerful tool due to its ability to detect lower frequency alleles. This study compared the variant detection capacity of NGS versus Sanger sequencing methods for resistance genotyping of 141 sequences (103 protease-reverse transcriptase tests and 38 integrase tests) in the pragmatic clinical laboratory setting of the University of Washington (UW) Clinical Virology Laboratory. NGS was found to be technically equivalent to Sanger for the detection of high frequency drug resistant variants. NGS detected more variants associated with higher levels of drug resistance in 18% of protease-reverse transcriptase sequences and 10% of integrase sequences, compared to Sanger, due to its capacity to detect low frequency variants. Clinical follow-up of sixty-nine patients for almost two years found some evidence to suggest that low frequency drug-resistant variants increase patients' risk of unsuppressed plasma HIV RNA (RR = 1.56; 95% CI: 0.75-3.22) and indicates the need for a larger study to confirm the association.

Introduction:
Antiretroviral drug treatment for HIV have been tremendously successful in reducing HIV morbidity, mortality, and transmission. However, rising rates of HIV infections resistant to antiretroviral (ARV) treatment pose a major threat to ongoing efforts to control the pandemic. [1] Routine HIV drug resistance genotyping improves health outcomes in HIV patients [2, 3] and is a cost-effective tool in the clinical management of HIV infection. [4]

Genotypic resistance testing assays have traditionally been based on Sanger sequencing, but interest in next-generation sequencing (NGS) technology is growing. While Sanger methods are highly reproducible and interpretable, minority HIV variants present in less than 20% of the viral population may escape detection (with a range of 10-30%, depending on the sample context). [5, 6] The capacity of NGS to provide additional data on low-frequency drug-resistant variants [7-10] and potential for lower costs per sample with large batches has led many clinical laboratories to consider transitioning from Sanger sequencing to NGS. [11] Previous research suggests a much higher prevalence of drug resistant-associated variants when a 1% detection of alleles is used compared to a standard 20%. [7, 12-14] However, whether the detection of low frequency variants has any effect on virological failure rates or other clinical outcomes remains unclear. Previous studies of low frequency resistance variants during integrase and protease inhibitor-based treatments have generally failed to find an association with virological failure. [15-20] In contrast, a pooled analysis of HIV-1 resistance variants specifically involving NNRTI-resistance, found a dose-dependent increased risk of virological failure with first-line ART. [21] Many previously published studies were conducted as part of drug or other clinical trials, thus lacking representation of real-world patient samples submitted for genotyping as part of routine care in a health care setting with reliable availability of multiple drug regimens.

The University of Washington Department of Laboratory Medicine, Virology Division, provides HIV resistance genotyping services for patients in the UW Medicine health system in Seattle, Washington, and surrounding areas. In 2016, the Clinical Virology Laboratory processed 925 clinician-ordered HIV drug resistance genotyping tests: 604 protease-reverse transcriptase (Pr-RT) and 321 integrase (INT) tests. These are standard tests, especially the protease-reverse transcriptase genotyping test, for patients entering care at UW Medicine and for subsequent clinical management of patients with persistently high plasma HIV RNA despite therapy. This study had two objectives: (1) compare the technical capacity of Sanger versus NGS sequencing methods for HIV resistance genotyping within the pragmatic patient population of UW Medicine, using two easily implementable NGS analysis pipeline for detection of drug resistance-associated variants (DRVs) and (2) assess the clinical impact of low frequency DRVs, which are undetectable by Sanger methods, on viral suppression over time.

Methods

Study Design
We designed an observational study to compare ability of two sequencing methods, NGS and Sanger, to detect HIV resistance genotyping in a real-world clinical setting of a large health system. We included 167 (119 Pr-RT and 48 INT) tests for which first-round amplicons were available from resistance genotyping assays processed by UW Clinical Virology Laboratory per standard clinical procedures between February 2 and May 24, 2016. To ensure genetic material was present in the sample, each test was confirmed to have detectable PCR products after a second round of amplification in the Sanger procedure for inclusion in the study. Sequences with <20,000 average reads were excluded to ensure sufficient coverage for detection of low frequency variants.

We used a retrospective cohort design to investigate whether the presence of low frequency variants was associated with subsequent viral suppression. Patients who received their medical care outside of UW Medicine were excluded as we lacked access to their records.

Data sources

Sanger sequences, variants, and interpretations: Results of previous genotypic resistance assays performed by the Clinical Virology Laboratory were obtained as the standard of care comparison group. HIV samples had been collected as part of routine care in UW Medicine and processed through the laboratory's standardized HIV genotypic resistance assay protocol, which involves two PCR amplification steps and Sanger sequencing. The sequence was interpreted for drug resistance variants by a trained laboratory technician. For variant calling, a mixture was defined as a position having a secondary peak at least 20% of the area under the curve (~20% allele frequency). Results were recorded in a Genotypic Resistance Assay Report that was returned to the patient's care provider via the electronic medical record. The report includes a list of variants associated with drug resistance in the three *pol* gene segments (Pr, RT, and INT) and an interpretation of the resistance profile (what drugs should be avoided) based on the detected variants. Variants listed in the Genotypic Resistance Assay Report were also interpreted using the Stanford HIV Drug Resistance Database (HIVdb version 8.5). [22] Resistance was recorded categorically according to the HIVdb standard categories: (0) no evidence of resistance; (1) potential low-level resistance; (2) low level resistance; (3) intermediate resistance; and (4) high level resistance. These categories were used in subsequent analyses to estimate mean level of resistance and compare major differences in detected levels of resistance.

NGS sequences, variants, and interpretations: First-round amplicons for each previously conducted Sanger test were retrieved from a -80°C archive and cleaned using 1.0X Ampure beads, quantitated on a Qubit 3.0, and diluted to 1 ng/uL. Libraries were prepped using quarter-reactions of Nextera XT followed by 15 cycles of PCR amplification and sequenced to achieve between 50,000-100,000 reads per sample on an Illumina MiSeq using both 1x192 and 2x300 bp runs. Samples were sequenced in batches of 20-24 samples to minimize the possibility of index cross-talk on a Miseq Illumina platform. Sequences were automatically trimmed in Geneious [23] and uploaded to two online variant callers: Paseq (https://www.paseq.org/) and HyDRA (https://hydra.canada.ca/). These two variant callers were selected because they are highly-developed, free variant callers with user-friendly web interfaces, requiring minimal to no bioinformatics skills, and both provide robust, reproducible, and easy to interpret results. [24] They also both use the well-established Stanford HIV Drug Resistance Database (at time of use: HIVdb version 8.5) to provide resistance interpretations. A low frequency DRV was defined as a resistance-associated variant between 1% and 20% frequency, based on the default 1% minimum frequency needed for a variant to be considered in the drug resistance report produced by both HyDRA and Paseq. A DRV was considered high frequency if detected at greater than 20% frequency. A sensitivity analysis considered a 2% allele frequency threshold as well explored whether any discordance between callers is possibly attributable to lack of sensitivity to detect low frequency variants.

Patient characteristics and clinical record abstraction: Patient age, sex, and sample date (baseline) were determined based on patient demographics recorded on the Genotypic Resistance Assay Report for all patients included in the test cohort. Patient electronic medical records were reviewed for clinical outcomes. Patients lacking a recorded plasma HIV RNA test result between the sample date and May 25, 2018 were excluded from the patient follow-up cohort due to lack of follow-up (patients not receiving care at UW Medicine). Laboratory and clinical visit records for patients included in the follow-up cohort were abstracted using a structured data form designed based on a preliminary in-depth search of five records prior to full record search and abstraction.

The following information was recorded for each patient: date of first plasma HIV RNA test from a UW Medicine laboratory, plasma HIV RNA at date of sample (or nearest date), drug regimen prescribed by clinician during clinic visit subsequent to sample date (when lab results, both genotyping and plasma HIV RNA, were available to inform clinician actions), and date of most recent office visit where HIV care was received. Adherence (yes/no) was assessed based on clinician visit notes as to whether the patient reported being adherent to the drug regimen prescribed after baseline (time invariant). Treatment

experience was assessed as (naïve/experienced) based on whether a new diagnosis was noted in the clinical record and/or a note was included about lack of prior treatment.

The primary outcome of interest, viral control or suppression (yes/no), was determined based on plasma HIV RNA measurements. Control or suppression was defined as a plasma HIV RNA concentration that was below the assay's threshold of detection (<1.6 or not detectable) for the duration of follow-up. Prescribed drug regimens were extracted from the medical record, and then grouped by category (PI, NRTI, NNRTI, and INTI) for analyses. Low frequency variant exposure was categorized as present/absent, according to whether the resistance profile called by NGS was higher than the profile reported by Sanger in the clinical record, due to the contributions of low frequency variants to the resistance profile. A second measure of risk categorized exposure as present when a patient was prescribed a drug regimen which would have been contraindicated based on the resistance profiles found by NGS.

Statistical analyses
Agreement of variants called by each NGS caller (HyDRA and Paseq) was analyzed for concordance using a Kappa statistic and agreement of allele frequency calls were compared using an intraclass correlation coefficient. A two-sided significance level of 0.05 was used for all comparisons, including correlations and t-tests used to compare relative mean level of resistance interpretations between Sanger and NGS due to low frequency DRVs. The relative risk ratios for whether an individual with low frequency DRVs failed to achieve viral suppression was computed using the *fmsb* package and interrater reliability statistics were calculated using the *irr* package. Statistical analyses were conducted in R (version 3.4.3) through the RStudio interface (version 1.0.153). Ethical approval for this study was granted by the University of Washington Institutional Review Board (STUDY#: 00000408).

Results:

Test Cohort
One hundred forty-one tests were ultimately included in the analysis (103 Pr-RT and 38 INT) with sufficiently high NGS sequencing coverage. Twenty-nine patients had both Pr-RT and INT tests ordered for them during the sample period and one patient had a repeat Pr-RT test ordered in the sample period. Therefore, 111 patients were included in the analysis. Most (95%) had HIV subtypes but 5 patients had subtype C (contributing 6 tests) and one had subtype A (1 test). Descriptive statistics for the test cohort are provided in Table 1.

Resistance Detected in Test Cohort
At the standard of care 20% allele frequency threshold, 39 (28%) tests were positive for resistance to at least one drug by both Sanger and NGS methods. At a 1% threshold, 62 (44%) tests were positive for resistance by HyDRA and 60 (43%) by Paseq. NNRTI resistance was most common at both allele frequency thresholds; PI and INTI resistance were least common. The five most prevalent high frequency DRVs were K103N and V108I (NNRTI DRVs), as well as M184V, T215S, and K65R (NRTI DRVs). The most prevalent low frequency DRVs were PI DRVs—D30N and M46V—and NNRTI DRVs—V108I, G190E, and F227L.

Of 103 Pr-RT, Sanger detected 2 (2% prevalence) with PI DRVs, 18 (17%) with NRTI DRVs, and 27 (26%) with NNRTI DRVs. Among 38 INT tests, 1 (3%) had INTI DRVs. Compared to Sanger, Paseq detected 8 (10%) more samples with PI resistance; 5 (22%) more samples with NRTI resistance, 10 (36%) more samples with NNRTI resistance, and 3 (11%) more samples with INTI resistance. HyDRA found a similarly higher prevalence of resistance: 9 (11%) more samples with PI DRVs, 8 (25%) more with NRTI DRVs, 7 (33%) more with NNRTI DRVs, and 4 more (13%) with INTI DRVs. Comparing

mean resistance levels, however, showed no significant difference between Sanger and NGS resistance interpretations (Table 3) for the whole sample set.

Agreement of Sanger and NGS

Sanger and NGS showed almost perfect concordance in their detection of high frequency DRVs (Kappa = 0.99). Only one Pr-RT sequence was discordant. In that sequence, Sanger called a mixed L210CW variant, which is associated with high resistance to didanosine (ddI), abacavir (ABC) and tenofovir (TDF). NGS, however, called this variant L201C, which is not associated with resistance. Upon manual inspection of the NGS sequence using Geneious, it became clear that the Sanger read was a false positive due to the inability of Sanger to interpret linkage between two adjacent nucleotide bases. In this instance, the reference codon was TTG (Leucine) but the viral population contained a mix of the reference codon and TGC (C - cysteine). However, because Sanger reads are produced one base at a time, Sanger interpreted the codon as having a mix of Gs at both the second and third bases, which led to an interpretation of TGCs (C - cysteine) and TGGs (W – tryptophan).

Agreement of NGS callers HyDRA and Paseq

At a 20% allele frequency threshold, HyDRA and Paseq had perfect agreement (Kappa = 1.00) in terms of which variants they detected (dichotomous: present/absent). At a 1% allele frequency threshold, Paseq and HyDRA called DRVs with very low agreement (Kappa = 0.03). However, the concordance between allele frequencies (continuous: 0-100%) detected by HyDRA versus Paseq found a very high level of agreement (intraclass correlation coefficient = 0.99; 95% CI: 0.988-0.994). (Figure 2) Discordance was primarily driven by low frequency DRVs: 7 samples had low frequency DRVs called only by Paseq and 7 samples had low frequency DRVs called only by HyDRA (14 samples total). Each of these discordant calls were for variants with very low frequency (median: 1.1%). A sensitivity analysis using a 2% threshold to compare the variants detected by each caller found Paseq and HyDRA had almost perfect agreement (Kappa = 0.83). Upon manual inspection of variants between 1-2% allele frequency, the discrepancies were attributable to minute differences in the number of reads registered by each caller. For further comparisons of NGS and Sanger results, the callers were assumed to have performed sufficiently similarly and thus the variant and resistance profiles called by HyDRA were used to represent the NGS results for subsequent analyses comparing NGS to Sanger.

Patient Follow-up Cohort

Of 111 patients included in the test cohort, 69 received care at UW Medicine and had follow up data. Descriptive statistics for patients in the follow-up cohort are provided in Table 2, divided between those who had low frequency DRVs detected by NGS (exposed; n=26) and those who did not (unexposed; n=43). There were no significant demographic or clinical differences between the exposed and unexposed in the cohort. In the total cohort, 32 patients (46%) lacked viral suppression during the follow-up period, 16 (23%) patients changed their drug regimen in follow-up, and 12 patients (17%) received repeat Sanger resistance tests as part of regular care (all of which detected no change in resistance profile).

Interpretations of Patient Drug Resistance Profiles in Patient Follow-up Cohort

Of sixty-nine patients with clinical follow-up, 26 (38%) had low frequency variants detected by NGS that were not found by Sanger. Low frequency DRVs were most commonly associated with NNRTI resistance (13 patients), followed by NRTI (11 patients), PI (7 patients), and INTI (4 patients). Thirteen patients received drugs which would have been contraindicated based on the resistance profiles found by NGS. Seven of these samples had major differences in the calls between Sanger and NGS, meaning that Sanger called no resistance and NGS-detected low frequency variants indicated intermediate or high resistance to a particular drug. For the remaining samples, NGS-detected low frequency variants were associated with potential or low levels of resistance.

Of 13 patients with contraindicated drug regimens according to NGS, three patients became virally suppressed and two of them reported perfect adherence. In contrast, 10 patients had high plasma HIV RNA and reported lack of adherence. Of 56 individuals who either had no low frequency variants or were not prescribed a contraindicated drug regimen, 29 were virally suppressed, of which 25 reported perfect adherence. In contrast, 27 patients were unsuppressed and only three reported adherence.

The relative risk of having unsuppressed plasma HIV RNA over the course of follow-up, was slightly higher for patients with any low frequency allele variants at baseline compared to those without (RR = 1.21; 95% CI: 0.63-2.33), although the findings were not statistically significant. When patients were defined as exposed if they had low frequency DRVs specifically associated with resistance to the category of one of their prescribed drugs (e.g. they were prescribed tenofovir and had a low frequency NRTI-resistance DRV), risk of detectable plasma HIV RNA was also higher (RR = 1.56; 95% CI: 0.75-3.22) compared to patients without regimen-related low frequency DRVs, although it did not reach statistical significance.

Discussion

In our comparison of rates of DRV detection using NGS versus Sanger, we found NGS detected more variants associated with higher levels of drug resistance in 18% of Pr-RT sequences and 11% of INT sequences, due to its capacity to detect low frequency DRVs. Low frequency DRVs were most commonly associated with NNRTI resistance, even though NNRTI drugs were rarely included in drug regimens of patients in the follow-up cohort (7%). These variants may be residual from selective pressure of previous regimens, as the cohort was predominantly composed of treatment-experienced patients. A quarter of patients in the follow-up cohort were put on a drug regimen to which they would have been contraindicated based on resistance profiles found by NGS at baseline, and these patients were more than 50% more likely to have unsuppressed plasma HIV RNA compared to patients without low frequency DRVs associated with their subsequent drug regimen.

Although two recent studies have found NGS to estimate similarly higher levels of antiviral drug resistance compared to Sanger for both Pr-RT and INT assays, [9, 10] studies involving clinical follow-up have found mixed results in measuring the association between low frequency variants and virological failure. As a further limitation, these association studies were conducted in clinical trial cohorts or aimed at testing a specific drug effect (e.g. PI or NNRTI). [15-21] Here we have presented the first study to investigate the effect of low frequency variants on viral suppression in patients on multi-drug therapeutic regimens in a pragmatic retrospective cohort.

Our study also directly compared the technical capacity of the two NGS variant callers, Paseq and HyDRA, and demonstrated high concordance of detection for high frequency DRVs and allele frequency measures for all DRVs. However, the callers disagreed in their detection of variants with very low frequency (1-2%), despite both claiming to provide sufficient sensitivity to detect variants with 1% frequency in supporting documentation. [24, 25] Discrepant calls for very low frequency variants may be a meaningful constraint for the implementation of these technologies in a clinical setting, if DRVs in this 1-2% range are ultimately associated with clinical resistance. This technical limitation has been called into question by other authors [26] and requires further investigation to inform the ongoing development of these variant callers.

By evaluating the utility of NGS in the context of a UW Medicine patient population, we provided a real-world example of its potential to inform clinical care and explore factors that could affect its implementation. However, UW Medicine provides HIV care to a high volume of patients, has strong HIV support services, and makes a wide variety of available HIV drugs accessible to patients. Standard of care indicates that resistance testing be conducted for patients upon engagement to care or for treatment-

experienced patient failing HIV therapy to inform selection of an appropriate ARV drug. In such a high-resource clinical setting, any indication of resistance on a Sanger resistance assay report (even at low levels) is sufficient to rule out that drug for an ARV regimen because numerous alternative options are available. All DRVs causing any degree of change to the level of drug resistance interpreted by the HIVdb were therefore included in this analysis. In less resource-rich settings, however, "potential" and "low-level" resistance may need to be excluded from the analysis because their clinical significance is not well understood and may be overly conservative for prescribing practices in resource-limited settings.

Furthermore, many health care providers, including UW Medicine, have added integrase inhibitors as a standard of care addition to patients' regimens because of its high efficacy for viral suppression and high barrier to resistance. More than 90% of patients in our cohort were prescribed a regimen that included an integrase inhibitor and just two of those patients had a low frequency DRV associated with INTI resistance. However, despite the high prevalence of INTI prescriptions and low prevalence of INTI DRVs, less than half of patients were virally suppressed during follow-up and more than half of all patients reported periods of non-adherence to their medications. This suggests that poor adherence is likely a major factor affecting rates of viral suppression in the cohort, regardless of the prevalence of low frequency variants. This hypothesis is further supported by the infrequent ordering of a repeat resistance tests and changes to patients' drug regimens during the follow-up period, even though many patients were not virally suppressed. In this patient population, clinicians seemed to predominantly attribute poor suppression as being caused by poor adherence rather than undetected resistance.

The primary limitation of our study is its low power. Despite significant efforts exerted to obtain a sufficiently large cohort size for the analysis, more than a third of patients in the test cohort were excluded from the sample cohort due to lack of follow-up. The exclusion of these individuals is not a major concern because their loss to follow-up is a regular occurrence in the course of clinical management and remains representative of a pragmatic population receiving testing in a clinical lab. However, given the clinical and potential economic importance of this study's objectives and the magnitude of effect on viral suppression found, a long-term cohort study involving many more patients receiving HIV care through UW Medicine is needed.

Our findings demonstrate two online, free, and easy-to-use NGS variant callers have high concordance with Sanger in a pragmatic clinical setting and thus may be good candidates for implementation in a clinical laboratory analysis pipeline for HIV drug resistance. Ultimately, this study did not have sufficient power to definitively measure an association between low frequency DRVs and virological failure. In this patient population, it seems likely that poor adherence is the primary drive of poor suppression and informs clinician's decision making about prescribed drug regiment, rather than stealth resistance from low frequency DRVs. A larger study with the capacity to incorporate adherence as a time-varying confounder would strengthen this suggested association.

Table 1. Descriptive Statistics of Test Cohort

HIV Genotypic Resistance Tests	(n = 141)	
Patient age, median (IQR)[1]	40 (18)	
Men, n (%)*	21 (87)	
Protease-reverse transcriptase	103	
Integrase	38	
B HIV subtype, n (%)	134 (95)	
C subtype tests	6	
A subtype	1	
Positive for any resistance at 20% threshold, n (% of tests)	39 (28)	
PI	2 (2)	
NRTI	18 (17)	
NNRTI	27 (26)	
INTI	1 (3)	
Low frequency DRV detected, n (%)	HyDRA[2] 62 (60)	Paseq 60 (58)
PI	11 (11)	10 (10)
NRTI	26 (25)	23 (22)
NNRTI	34 (33)	37 (36)
INTI	5 (13)	4 (11)

[1] Tests collected from 111 unique patients: 29 patients had both Pr-RT and INT tests and 1 patient had a repeat Pr-RT test during the sample period.
(*Protease-reverse transcriptase test*) PI: Protease Inhibitors; NRTI: Nucleoside/Nucleotide Reverse Transcriptase Inhibitors; NNRTI: Non-Nucleoside Reverse Transcriptase Inhibitors
(*Integrase test*) INTI: Integrase Inhibitors
[2] HyDRA DRVs were used as the "NGS" value for subsequent analyses in the patient follow-up cohort that compared Sanger to NGS resistance profiles.

Table 2. Descriptive Statistics of Patient Follow-up Cohort

Patient Characteristics	Patients with clinical follow-up (n = 69)		p-value*
	No low frequency DRVs (n=43)	Low frequency DRVs detected (n=26)	
Patient age – median (IQR)	40 (30 – 49)	42 (36 -52)	0.16
Men (%)	40 (93)	21 (81)	0.25
Follow-up [days] – median (IQR)	656 (566-700)	711 (568-763)	0.60
Plasma HIV RNA at baseline [log10(copies/mL)] – median (IQR)	4.0 (3.6-5.1)	4.1 (3.9-4.6)	0.63
Missing plasma HIV RNA measure at baseline	0	1	-
Changed drug regimen from baseline (%)	13 (30)	4 (15)	0.27
Received repeat Sanger resistance genotyping in follow-up (%)	9 (21)	3 (12)	0.50
Adherent (%)	20 (47)	11 (42)	0.93
Treatment naïve (%) *Missing*	0 8	8 (31) 4	-
Drugs included in prescribed regimen (% of patients)			
PI	15 (35)	6 (23)	0.45
NRTI	43 (100)	26 (100)	-
NNRTI	3 (7)	0 (0)	0.44
INTI	36 (84)	22 (85)	1
Low frequency DRV detected (% of patients)			
PI	0	7 (27)	-
NRTI	0	11 (42)	-
NNRTI	0	13 (34)	-
INTI	0	4 (15)	-

Table 3. Mean levels of predicted reduced ARV susceptibility of the 103 Pr-RT and 38 INT sequences according to Sanger genotypic resistance testing versus NGS.

Drug class	Drug	Sanger (Mean ± SD)	NGS (Mean ± SD)	p-value*
NRTIs (n = 103 sequences)	3TC	0.54 ± 1.38	0.66 ± 1.47	0.56
	ABC	0.443 ± 1.10	0.54 ± 1.21	0.47
	AZT	0.19 ± 0.75	0.25 ± 0.79	0.59
	TDF	0.16 ± 9	0.29 ± 0.88	0.21
NNRTIs (n = 103 sequences)	EFV	0.88 ± 1.60	1.07 ± 1.67	0.42
	ETR	0.17 ± 0.62	0.34 ± 0.81	0.10
	RPV	0.25 ± 0.87	0.54 ± 1.19	0.05
PIs (n = 103 sequences)	ATV	0.02 ± 0.20	0.10 ± 0.36	0.05
	DRV	0	0	-
	LPV	0.03 ± 0.30	0.07 ± 0.38	0.41
INTIs (n = 103 sequences)	DTG	0.053 ± 0.23	0.11 ± 0.51	0.37
	EVG	0.26 ± 0.83	0.34 ± 0.99	0.22
	RAL	0.26 ± 0.83	0.34 ± 0.99	0.22

Footnote: Susceptible (0); Potential low-level resistance (1); Low level resistance (2); Intermediate resistance (3); High-level resistance (4). *Student's t-test.

Figure 1. Diagram of study cohort participants.

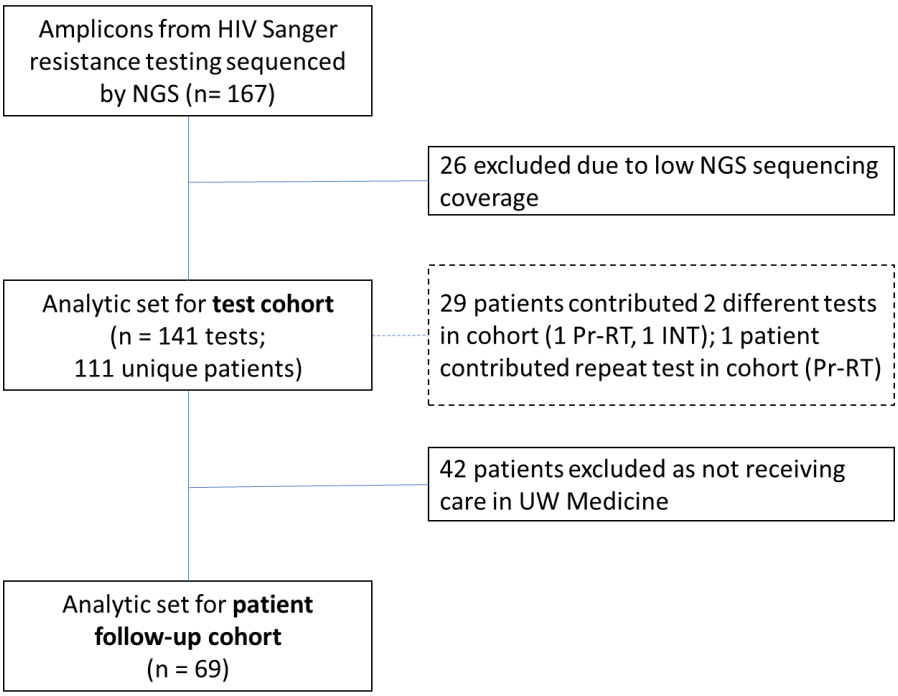
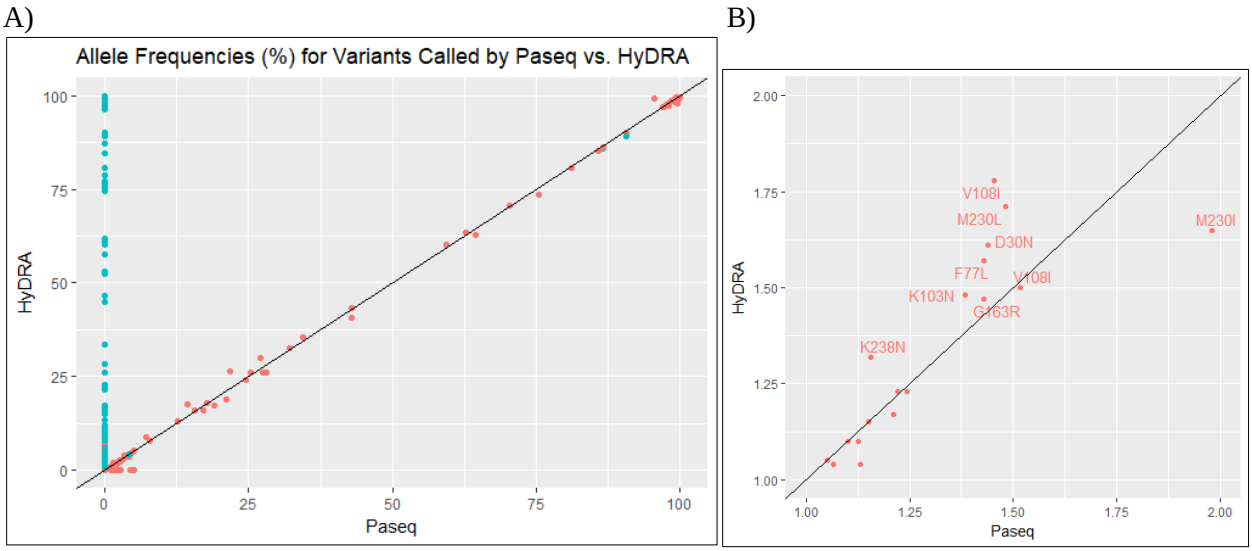


Figure 2. Concordance of allele frequency calls by HyDRA and Paseq variant callers.



(A) Paired allele frequency reads for variants called by Paseq and HyDRA. (B) Close-up focus on 1-2% frequency range (cluster of observations in bottom left corner of Figure S1-A). The default setting of HyDRA includes calls of many more accessory variants than Paseq, which led to the large number of accessory mutations called by HyDRA but not (allele frequency = 0) by Paseq. *Student's t-test.

Acknowledgements: Negar Makhsous for conducting all of the next-generation sequencing procedures and Alexander L. Greninger for facilitating my access to clinical samples and sequencing data and for providing mentorship throughout the project.

Bibliography

1. WHO. HIV drug resistance: tackling HIV drug resistance: trends, guidelines and global actions: policy brief. 2017. doi: http://www.who.int/iris/handle/10665/255881.
2. Baxter JD, Mayers DL, Wentworth DN, Neaton JD, Hoover ML, Winters MA, et al. A randomized study of antiretroviral management based on plasma genotypic antiretroviral resistance testing in patients failing therapy. CPCRA 046 Study Team for the Terry Beirn Community Programs for Clinical Research on AIDS. Aids. 2000;14(9):F83-93. Epub 2000/07/14. PubMed PMID: 10894268.
3. Palella FJ, Jr., Armon C, Buchacz K, Cole SR, Chmiel JS, Novak RM, et al. The association of HIV susceptibility testing with survival among HIV-infected patients receiving antiretroviral therapy: a cohort study. Ann Intern Med. 2009;151(2):73-84. Epub 2009/07/22. PubMed PMID: 19620160.
4. Sax PE, Islam R, Walensky RP, Losina E, Weinstein MC, Goldie SJ, et al. Should resistance testing be performed for treatment-naive HIV-infected patients? A cost-effectiveness analysis. Clin Infect Dis. 2005;41(9):1316-23. Epub 2005/10/06. doi: 10.1086/496984. PubMed PMID: 16206108.
5. Leitner T, Halapi E, Scarlatti G, Rossi P, Albert J, Fenyo EM, et al. Analysis of heterogeneous viral populations by direct DNA sequencing. Biotechniques. 1993;15(1):120-7. Epub 1993/07/01. PubMed PMID: 8363827.
6. Halvas EK, Aldrovandi GM, Balfe P, Beck IA, Boltz VF, Coffin JM, et al. Blinded, Multicenter Comparison of Methods To Detect a Drug-Resistant Mutant of Human Immunodeficiency Virus Type 1 at Low Frequency. J Clin Microbiol. 2006;44(7):2612-4. doi: 10.1128/jcm.00449-06. PubMed PMID: 16825395.
7. Cozzi-Lepri A, Noguera-Julian M, Di Giallonardo F, Schuurman R, Daumer M, Aitken S, 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. 2015;70(3):930-40. Epub 2014/10/23. doi: 10.1093/jac/dku426. PubMed PMID: 25336166; PubMed Central PMCID: PMCPMC4319483.
8. Moscona A, Peluso RW. Fusion properties of cells persistently infected with human parainfluenza virus type 3: participation of hemagglutinin-neuraminidase in membrane fusion. J Virol. 1991;65(6):2773-7. Epub 1991/06/01. PubMed PMID: 1851852; PubMed Central PMCID: PMCPMC240891.
9. Tzou PL, Ariyaratne P, Varghese V, Lee C, Rakhmanaliev E, Villy C, et al. Comparison of an In Vitro Diagnostic Next-Generation Sequencing Assay with Sanger Sequencing for HIV-1 Genotypic Resistance Testing. J Clin Microbiol. 2018;56(6). Epub 2018/04/06. doi: 10.1128/jcm.00105-18. PubMed PMID: 29618499; PubMed Central PMCID: PMCPMC5971553.
10. Dunn DT, Stöhr W, Arenas-Pinto A, Tostevin A, Mbisa JL, Paton NI, et al. Next generation sequencing of HIV-1 protease in the PIVOT trial of protease inhibitor monotherapy. J Clin Virol. 2018;101:63-5. doi: 10.1016/j.jcv.2018.02.003. PubMed PMID: PMC5861306.
11. Dudley DM, Chin EN, Bimber BN, Sanabani SS, Tarosso LF, Costa PR, et al. Low-cost ultra-wide genotyping using Roche/454 pyrosequencing for surveillance of HIV drug resistance. PLoS One. 2012;7(5):e36494. Epub 2012/05/11. doi: 10.1371/journal.pone.0036494. PubMed PMID: 22574170; PubMed Central PMCID: PMCPMC3344889.
12. Todesco E, Charpentier C, Bertine M, Wirden M, Storto A, Desire N, et al. Disparities in HIV-1 transmitted drug resistance detected by ultradeep sequencing between men who have sex with men and heterosexual populations. HIV Med. 2017;18(9):696-700. Epub 2017/04/27. doi: 10.1111/hiv.12508. PubMed PMID: 28444829.
13. Storto A, Visseaux B, Bertine M, Le Hingrat Q, Collin G, Damond F, et al. Minority resistant variants are also present in HIV-2-infected antiretroviral-naive patients. J Antimicrob Chemother. 2018. Epub 2018/02/08. doi: 10.1093/jac/dkx530. PubMed PMID: 29415189.
14. Johnson JA, Li JF, Wei X, Lipscomb J, Irlbeck D, Craig C, et al. Minority HIV-1 drug resistance mutations are present in antiretroviral treatment-naive populations and associate with reduced treatment efficacy. PLoS Med. 2008;5(7):e158. Epub 2008/08/01. doi: 10.1371/journal.pmed.0050158. PubMed PMID: 18666824; PubMed Central PMCID: PMCPMC2488194.

15. Perrier M, Visseaux B, Landman R, Joly V, Todesco E, Yazdanpanah Y, et al. No impact of HIV-1 protease minority resistant variants on the virological response to a first-line PI-based regimen containing darunavir or atazanavir. J Antimicrob Chemother. 2018;73(1):173-6. Epub 2017/10/28. doi: 10.1093/jac/dkx366. PubMed PMID: 29077926.
16. Charpentier C, Laureillard D, Piketty C, Tisserand P, Batisse D, Karmochkine M, et al. High frequency of integrase Q148R minority variants in HIV-infected patients naive of integrase inhibitors. Aids. 2010;24(6):867-73. Epub 2010/02/18. doi: 10.1097/QAD.0b013e3283367796. PubMed PMID: 20160635.
17. Ross LL, Weinberg WG, DeJesus E, Fischl MA, Horton JH, Pappa KA, et al. Impact of low abundance HIV variants on response to ritonavir-boosted atazanavir or fosamprenavir given once daily with tenofovir/emtricitabine in antiretroviral-naive HIV-infected patients. AIDS Res Hum Retroviruses. 2010;26(4):407-17. Epub 2010/04/13. doi: 10.1089/aid.2009.0189. PubMed PMID: 20380480; PubMed Central PMCID: PMCPMC2875981.
18. Metzner KJ, Rauch P, von Wyl V, Leemann C, Grube C, Kuster H, et al. Efficient Suppression of Minority Drug-Resistant HIV Type 1 (HIV-1) Variants Present at Primary HIV-1 Infection by Ritonavir-Boosted Protease Inhibitor-Containing Antiretroviral Therapy. J Infec Dis. 2010;201(7):1063-71. doi: 10.1086/651136.
19. Lataillade M, Chiarella J, Yang R, Schnittman S, Wirtz V, Uy J, 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. Epub 2010/06/10. doi: 10.1371/journal.pone.0010952. PubMed PMID: 20532178; PubMed Central PMCID: PMCPMC2880604.
20. Simen BB, Simons JF, Hullsiek KH, Novak RM, Macarthur RD, Baxter JD, et al. Low-abundance drug-resistant viral variants in chronically HIV-infected, antiretroviral treatment-naive patients significantly impact treatment outcomes. J Infect Dis. 2009;199(5):693-701. Epub 2009/02/13. doi: 10.1086/596736. PubMed PMID: 19210162.
21. Li JZ, Paredes R, Ribaudo H, Svarovskaia ES, Metzner KJ, Kozal MJ, et al. Minority HIV-1 Drug Resistance Mutations and the Risk of NNRTI-based Antiretroviral Treatment Failure: A Systematic Review and Pooled Analysis. Jama. 2011;305(13):1327-35. doi: 10.1001/jama.2011.375. PubMed PMID: PMC3325645.
22. Liu TF, Shafer RW. Web resources for HIV type 1 genotypic-resistance test interpretation. Clin Infect Dis. 2006;42(11):1608-18. Epub 2006/05/03. doi: 10.1086/503914. PubMed PMID: 16652319; PubMed Central PMCID: PMCPMC2547473.
23. Kearse M, Moir R, Wilson A, Stones-Havas S, Cheung M, Sturrock S, et al. Geneious Basic: an integrated and extendable desktop software platform for the organization and analysis of sequence data. Bioinformatics. 2012;28(12):1647-9. Epub 2012/05/01. doi: 10.1093/bioinformatics/bts199. PubMed PMID: 22543367; PubMed Central PMCID: PMCPMC3371832.
24. Noguera-Julian M, Edgil D, Harrigan PR, Sandstrom P, Godfrey C, Paredes R. Next-Generation Human Immunodeficiency Virus Sequencing for Patient Management and Drug Resistance Surveillance. J Infect Dis. 2017;216(suppl_9):S829-s33. Epub 2017/10/03. doi: 10.1093/infdis/jix397. PubMed PMID: 28968834; PubMed Central PMCID: PMCPMC5853595.
25. Nykoluk M, Taylor T, Ji H, Enns E. HyDRA Web User Guide. 2016.
26. Howison M, Coetzer M, Kantor R. Measurement error and variant-calling in deep Illumina sequencing of HIV. bioRxiv. 2018.