Validation of a high throughput next generation sequencing assay for HIV drug resistance genotyping

Du-Ping Zheng[1]*, Josh DeVos[1], Kevin Tang[2], Heather Alexander[1], Guoqing Zhang[1], Clement Zeh[1]

1.International Laboratory Branch, DGHT/CGH, 2. Biotechnology Core Facility Branch, DSR/NCEZID, US Centers for Disease Control and Prevention. Atlanta, USA

Background

Globally, the number of people living with HIV on antiretroviral treatment (ART) reached 28.7 million in 2021. To ensure the effectiveness of treatment regimens, WHO issued guidelines and action plans for systematic surveillance of HIV drug resistance (DR), including monitoring HIVDR to dolutegravir. Currently, Sanger-based sequencing is the primary technology for HIVDR detection and surveillance. However, it has limitations, including low throughput, high cost, and less sensitivity in detecting variants below 20%. In this study, we developed an amplicon-based next generation sequencing (NGS) using illumina technology on MiSeq Platform for HIVDR genotyping and assessed its accuracy, precision, reproducibility, and sensitivity compared to the Sanger sequencing method.

Methods

PCR amplicons of HIV protease, reverse transcriptase (PRT), and integrase (INT) genes from 48 analytic samples representing 8 major subtypes and recombinants (>89% of all HIVs) were generated using ThermoFisher HIVDR genotyping kit and sequenced with illumina Nextera-XT kit. NGS sequences were compared with Sanger sequences and analyzed statistically to assess accuracy, precision, reproducibility, and variants detection at 10%, 15%, and 20% thresholds.

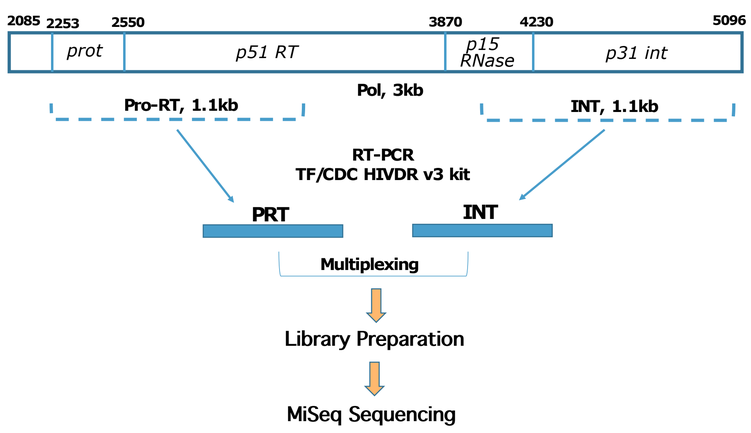
Results

Both PRT and INT NGS sequences from 48 amplicon samples exhibited overall >99.5% accuracy (CI 99.5-100%) compared to Sanger sequences. For detecting HIVDR mutations, NGS had 99.7% agreement in PRT and 100% in INT with Sanger sequencing. For precision, NGS produced an overall 100% (ranged 99.8-100%) similarity within 8 replicates from each 12 samples. With the same 8 replicates of 12 samples, NGS generated almost identical data in PRT (99.6%) and INT (99.9%) between 3 independent runs (p=1). In a 96-sample run, NGS generated an average of 30mb data and 33,157 reads coverage per sample, which is sufficient for variant calls. In this sample panel, NGS detected an average of 2.15 and 4.77 more variants in PRT, 1.85 and 4.35 more in INT at 15% and 10% threshold compared to the Sanger sequencing at 20% threshold, respectively.

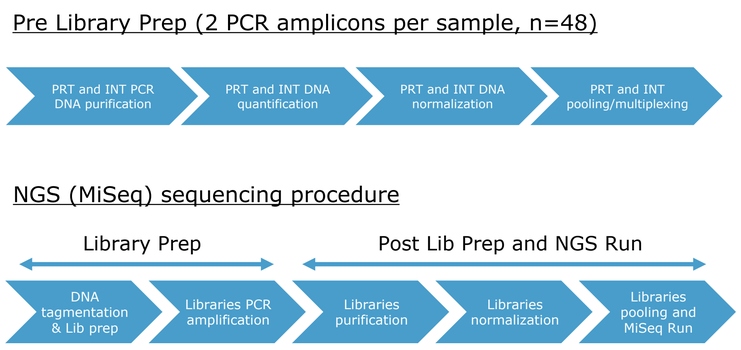
Conclusion

We successfully validated illumina-based NGS for HIVDR genotyping with high accuracy and precision compared to Sanger sequencing. The validated NGS provides a higher throughput, potentially lower cost at scale, and sensitive sequencing method for detecting a full spectrum of HIVDR mutations, which can strengthen current HIVDR surveillance and preserve and guide effective ART regimens.

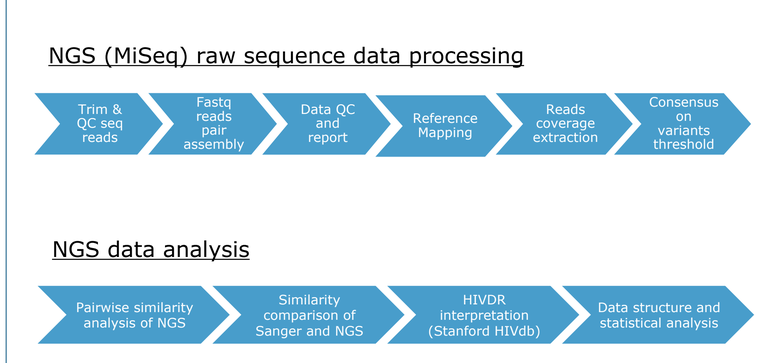
Overview Lab Workflow of NGS for HIVDR Detection



NGS Wet Lab Procedures (after optimization)



NGS Dry Lab Data Processing and Analysis



Validation Criteria and Plan

WHO recommended validation criteria				
Test samples	>=20 samples	Major subtypes	HIVDR mutations	
Assess criteria	Accuracy (n=20x1)	Precision (n=3x5)	Reproducibility (n=3x5)	Sensitivity (Amp, TF validation)
Accept criteria	≥ 90% of pairwise comparisons for each sample must be ≥ 98% identical			

Validation Assessing Criteria	Samples	Replicates	No. of tests	Test method	Subtotal
Accuracy	48	2	1	Sanger + NGS	96
Precision (intra)	12	8	3	NGS	288
Reproducibility (inter)	12	8	3	NGS	288
Sensitivity (10, 15, 20%)	48	1	1	NGS	48

Attributes of Samples Used for NGS Validation (n=48)

Viral Load (copies/mL)

1000- 5000	5000-10000	10000-100000	>100000	unknown
14	13	16	1	4

Subtypes (n=8)

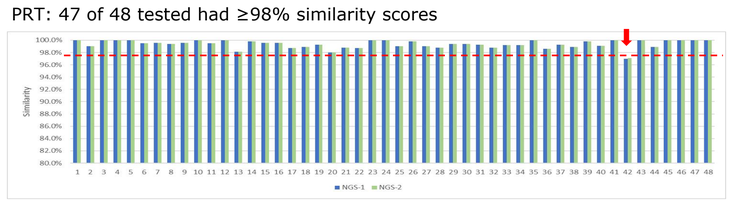
A (A1)	B	C	D	F (F1)	01_AE	02_AG	02_AG/A1
2	23	9	3	4	2	2	3

Sample type

Plasma	DBS	Clone
34	10	4

Accuracy: Pairwise Similarity between Sanger and NGS (n=48, 2 replicates, 1 run)

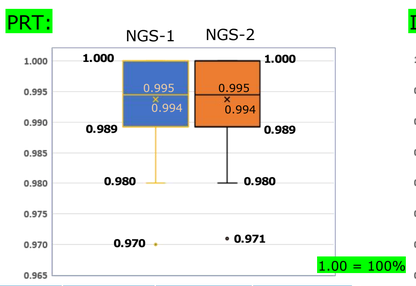
PRT: 47 of 48 tested had ≥98% similarity scores



INT: 48 of 48 tested had ≥98% similarity scores

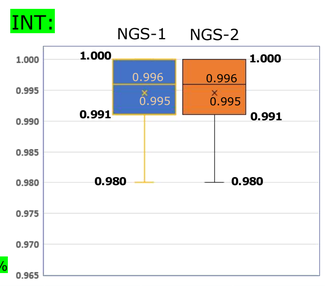
Sanger and NGS Identity Scores of NGS-1 & -2 and between

PRT:



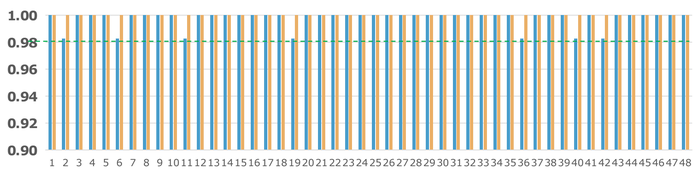
PRT	NGS-1	NGS-2	NGS1:NGS2
Mean	0.99	0.99	1.00
Std Dev	0.00645	0.00637	0.00039
CI (95%)	0.99 – 1.00	0.99 – 1.00	1.00
p (T. Test)			0.99

INT:

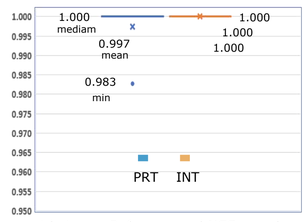


INT	NGS-1	NGS-2	NGS1:NGS2
Mean	0.99	0.99	1.00
Std Dev	0.00552	0.00552	0.00040
CI (95%)	0.99 – 1.00	0.99 – 1.00	1.00
p (T. Test)			1.00

Accuracy of HIV DR Mutations Detected by NGS (n=48)



DRMs	PRT	INT
Mean	0.997	1.000
Median	1.000	1.000
Mode	1.000	1.000
Standard Deviation	0.006	0.000
Sample Variance	0.000	0.000
Minimum	**0.983**	1.000
Maximum	1.000	1.000



Assessed at all DR sites of PRT (n=58) and INT (n=22) for agreement between Reference and NGS samples

Precision and Reproducibility Test Results (N=12, 8 replicates, Protease and Reverse Transcriptase, PRT)

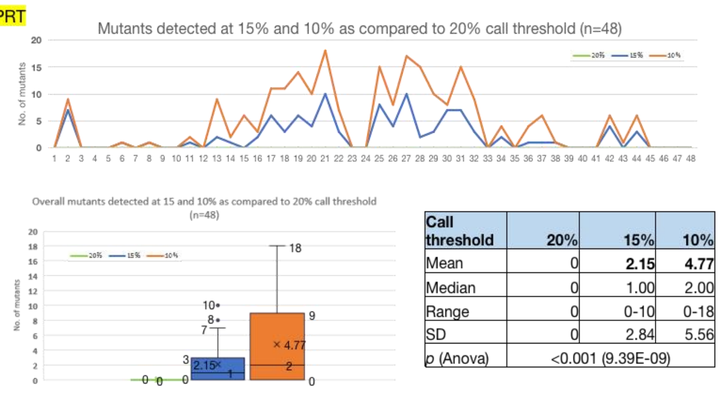
PRT 8 replicates	PRS04			PRS05			PRS06			PRS08		
	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3
Minimum	1	1	1	1	1	1	1	1	1	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	1	1	1	1	1	1	1	1	1	1	1
Std Dev	0	0	0	0	0	0	0	0	0	0	0	0
P-value (anova)	1			1			1			1		
	PRS10			PRS13			PRS14			PRS16		
Minimum	1	1	1	1	1	1	1	1	1	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	1	1	1	1	1	1	1	1	1	1	1
Std Dev	0	0	0	0	0	0	0	0	0	0	0	0
P-value (anova)	1			1			1			1		
	PRS26			PRS28			PRS29			PRS43		
Minimum	1	1	1	1	0.999	0.999	1	1	1	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	1	1	1	0.999429	0.999464	1	1	1	1	1	1
Std Dev	0	0	0	0	0.000504	0.000508	0	0	0	0	0	0
P-value (anova)	1			0.999999633			1			1		

Precision and Reproducibility Test Results (n=12, 8 replicates, Integrase, INT)

INT 8 replicates	PRS04			PRS05			PRS06			PRS08		
	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3	Run-1	Run-2	Run-3
Minimum	0.998	1	0.998	1	1	0.996	1	1	1	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	0.9995	1	0.9995	1	1	0.998286	1	1	1	1	1	1
Std Dev	0.000882	0	0.000882	0	0	0.002016	0	0	0	0	0	0
P-value (anova)	0.999999125			0.999995429			1			1		
	PRS10			PRS13			PRS14			PRS16		
Minimum	1	1	1	1	1	1	1	1	0.998	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	1	1	1	1	1	1	1	0.999143	1	1	1
Std Dev	0	0	0	0	0	0	0	0	0.001008	0	0	0
P-value (anova)	1			1			0.999998859			1		
	PRS26			PRS28			PRS29			PRS43		
Minimum	1	1	1	0.998	1	1	1	1	0.997	1	1	1
Maximum	1	1	1	1	1	1	1	1	1	1	1	1
Mean	1	1	1	0.9995	1	1	1	1	0.998607	1	1	1
Std Dev	0	0	0	0.000882	0	0	0	0	0.001197	0	0	0
P-value (anova)	1			0.999999126			0.999996589			1		

Sensitivity Assessment on Viral Variants Detection

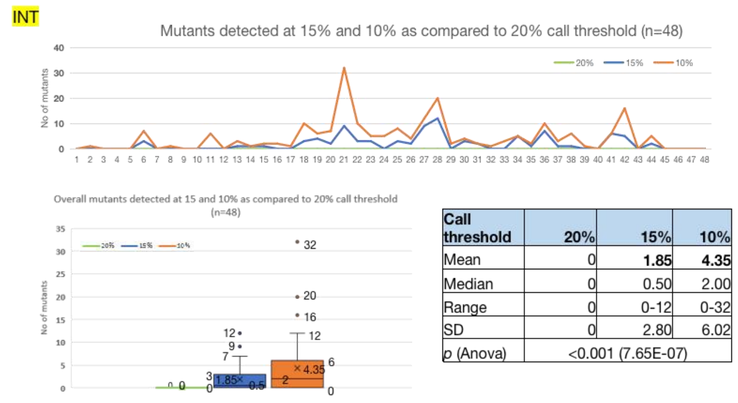
PRT



Call threshold	20%	15%	10%
Mean	0	**2.15**	**4.77**
Median	0	1.00	2.00
Range	0	0-10	0-18
SD	0	2.84	5.56
p (Anova)	<0.001 (9.39E-09)		

Sensitivity Assessment on Viral Variants Detection

INT



Call threshold	20%	15%	10%
Mean	0	**1.85**	**4.35**
Median	0	0.50	2.00
Range	0	0-12	0-32
SD	0	2.80	6.02
p (Anova)	<0.001 (7.65E-07)		

Presented at IAS 2023, the 12th IAS Conference on HIV Science

IAS 2023 23 – 26 July