

A new sensitive and robust next-generation sequencing platform for HIV-1 drug resistance mutations testing

Bin Yu^{*1}, Changzhong Jin^{*2}, Zixuan Ma¹, Ziwei Cai¹, Tingsen Li¹, Dan Wang¹,
Wenwen Xiao¹, Yanghao Zheng², Wanpeng Yin², Nanping Wu², and Miao Jiang¹

¹Beijing Liangxin Biotechnology Development LTD., 5/F, Unit 3, Building 6, Courtyard No. 88, Scientific Innovation Street 6, Beijing Technical Economic Development Area, Beijing, China

²State Key Laboratory for Diagnosis and Treatment of Infectious Diseases, National Clinical Research Center for Infectious Disease, National Medical Center for Infectious Disease, The First Affiliated Hospital, School of Medicine, Zhejiang University, Hangzhou 310003, China.

*Equal contribution

Email:

Yubin@boxinbio.com (Bin Yu);

changzhongjin@163.com (Jin Changzhong) ;

jiangmiao@boxinbio.com (Miao Jiang);

flwnp2013@163.com (Nanping Wu)

Abstract

Next-generation sequencing (NGS) is a trending new platform which allows cheap, quantitative, high-throughput, parallel sequencing for minority variants with frequencies less than 20% of the HIV-1 quasi-species. In clinical setting, these advantages are crucial for choosing antiretroviral drugs with low genetic barriers and will potentially benefit treatment outcomes.

In this investigation, we implemented the Boxin HIV-1 NGS platform for genotyping the drug-resistance-associated variants in PR/RT regions. Plasmids with known mutations were used to analyze the accuracy, reproducibility, and reliability of the Boxin NGS assay. Variant frequencies

reported by Boxin NGS and the theoretical value were highly concordant. The Bland-Altman plot and the coefficient of variation (7%) suggested that the method has excellent reproducibility and reliability. Sanger sequencing confirmed the existence of these known variants with frequencies equal or above 20%.

78 blood samples were obtained from AIDS patients and underwent PR/RT region genotyping by Sanger sequencing and Boxin NGS. 33 additional drug resistance mutations were identified by Boxin NGS, 23/33 mutations were minority variants with frequencies below 20%.

15 blood samples obtained from AIDS patients underwent PR/RT region genotyping by Sanger sequencing, Boxin NGS, and Vela NGS. The Bland-Altman plot suggested that the variant frequencies detected by Boxin and Vela were highly concordant. Moreover, Boxin NGS assay detected five more minority variants with frequencies ranged from 1% to 20%. In a series of samples collected from 2016 to 2017, Boxin NGS reported a M184V mutation with a frequency of 4.92%, 3 months earlier than this mutation was firstly detected by Vela NGS and Sanger sequencing.

In conclusion, Boxin NGS had good accuracy, reproducibility, and reliability. Boxin NGS was highly concordant with Sanger sequencing and Vela NGS. In terms of genotyping HIV-1 variants in PR/RT regions, Boxin NGS was more cost-efficient and appeared to have increased sensitivity without compromising sequence accuracy.

Introduction

HIV-1 drug resistance (HIVDR) is the primary challenge to administrate an effective antiretroviral therapy (ART)¹⁻⁴. Error-prone and rapid viral replication result in an extremely diversified HIV-1 quasi-species^{3,5-7}. Under selection pressure, gene mutation-induced low-frequency drug resistance can surpass the effect of ART, become the dominant species in the host, give rise to a secondary resistance, and lead to a worse prognosis or a virological failure^{2,4-6,8-11}. Therefore, identifying low-frequency HIVDRs is crucial for administering a successful ART^{2,12-15}.

Sanger sequencing was considered as the gold standard for detecting drug-resistance mutations (DRMs)¹⁶. However, its inability of detecting low abundance drug-resistance mutations (LADRM) present in less than 20% of the viral quasi-species is considered as a major obstacle^{1,4,5,9,16,17}.

Compared to Sanger sequencing, NGS guarantees a cheaper, sensitive, parallel, and high-throughput genome sequencing method^{4,16,18}. NGS allows a lower limit of detection of 1%, enhances the sequencing resolution, and deconvolutes complex HIV-1 quasi-species^{5,6,19}. The ability of NGS-based genotyping in providing numerical data allows it to report the relative abundance of the variants and predict DRMs^{5,17,20,21}. Recent studies have demonstrated the importance of identifying patients who are prone to virological failure due to an increasing evidence of LADRM inducing immune system suppression, drug resistance aggregation, and death^{5,22-24}.

In this study, we implemented and evaluated the Boxin HIV-1 NGS platform (Liangxin Biotechnology Development LTD) in terms of its accuracy, reproducibility, and reliability. We also compared the performance of Sanger sequencing, the Vela Diagnostics Sentosa SQ HIV genotyping NGS assay, and the Boxin NGS HIV-1 genotyping assay in testing PR/RT sequences.

Methods

Ethical statement

This study was reviewed and waived by the Clinical Research Ethics Committee of the First Affiliated Hospital, College of Medicine, Zhejiang University.

Evaluation of the Boxin NGS using plasmids

Accuracy, reproducibility, and reliability of the Boxin NGS assay were investigated by genotyping amino acid variants in plasmids with known mutations. The plasmids also underwent Sanger sequencing to verify the existence of the variants. Three different HIV-1 plasmids were obtained from Sangon Biotech LTD, Shanghai, China. The plasmids were diluted to 0.2ng/ μ L, and the wild type plasmid(YB002) was mixed with the plasmid with a single variant (YB003) or the plasmid with 7 variants (YB012) as shown in table 1. Each sample was prepared examined in triplicates, and the procedure was repeated twice by two different lab technicians who worked independently. Each sample was amplified by MiniAmp PCR (ThermoFisher). Sanger sequencing was performed by a reference laboratory (BGI Genomics, Beijing, China).

Patients and samples

87 blood samples were collected from AIDS patients treated at the First Affiliated Hospital, College of Medicine, Zhejiang University. All blood samples underwent Boxin NGS and Sanger sequencing. Another 15 samples were obtained from AIDS patients who were treated at the Beijing Ditan Hospital Capital Medical University, and underwent Sanger sequencing, Boxin NGS, and Vela NGS. Reverse transcriptase (RT) and protease (PR) sequences were genotyped. Protease inhibitors (PIs), nucleoside reverse transcriptase inhibitors (NRTIs), and nonnucleoside reverse transcriptase inhibitors (NNRTIs) associated DRMs were analyzed and reported by all three methods. Integrase inhibitors (INs) were not analyzed for the purpose of the study.

Boxin NGS

Ion Plus Fragment Library Kit (Thermo Fisher Scientific, Waltham, USA) and Ion ShearTM Plus Reagents Kit (Thermo Fisher Scientific) were the reagents employed for library construction and DNA shearing, respectively.

RNA was extracted from 500 μ L of each blood sample. Ion Torrent's instruments (Thermo Fisher Scientific) were employed for deep sequencing. Two wild-type HIV-1 B subtype plasmids (positive and negative) were added to each run and was run in parallel with other samples as controls. 96 samples were examined for each run, including a positive and a negative control. RT-PCR was conducted to amplify the sample using the MiniAmp PCR Instrument (ThermoFisher). Library was by Ion Chef Instrument (ThermoFisher). A quantitative PCR was carried out to quantify the library. The bioinformatics pipeline was developed by ThermoFisher Scientific with optimization.

Sanger sequencing and Vela NGS

500 μ L of each blood sample was sent away to reference laboratories. Vela NGS assay was performed as per manufacturer's instructions at Beijing Ditan Hospital Capital Medical University. Sanger sequencing was performed at a reference library (Tsingke Biotechnology Co., Ltd).

Database and data analysis

Results were analyzed and reported using the NGS database developed by Liangxin (Beijing) Biotechnology Development LTD. The NGS database was constructed and optimized based on populated Chinese AIDS patients, which allowed identification and analysis of 256 PI, NRTI, NNRTI, and IN associated DRMs. Figures and tables were generated by Microsoft Excels.

Results

Evaluation of the Boxin NGS assay using plasmids

Regression analysis was applied to determine whether the variant frequencies reported by Boxin NGS and the theoretical values were concordant. The theoretical values of variant frequencies were graphed against the mean experimental results. The experimental data distributed close to the predicted values. The slope of the line was 0.8948, and the y-intercept was -0.7312. The coefficient of determination R^2 was 0.9932 (*Fig. 1A*).

The reproducibility and reliability of the Boxin NGS assay were analyzed by comparing the mutation frequencies detected and reported by two lab technicians. Each plasmid sample was made in triplicates and the experiment was repeated twice by two examiners. The average value of the experimental variant frequency of each sample was reported by each examiner. The reproducibility was visualized by the Bland-Altman plot (*Fig 1B*); the average variant frequency of the same sample was graphed against the difference between the variant frequencies reported by two examiners of the same sample. All data points distributed within the limits of agreement and closely to the line of bias. The line of bias was the average of the difference, which was -0.908. The lines of upper and lower limit of agreement were calculated as bias \pm 1.96*STDEV, which was found to be 2.504 and -4.321, respectively.

The reliability of repeated measurements could be determined by the coefficient of variation. The coefficient of variation was calculated as the standard deviation divided by the mean variant frequencies reported by two examiners. The average coefficient of variation was found to be 7%.

Sanger sequencing assay was carried out to verify the mutations in the 10 samples constructed using plasmids. When the samples were prepared with variant frequencies equal and above 20%, Boxin NGS and Sanger sequencing reported the same mutations (*Fig. 1C*). When the samples were prepared with variant frequencies below 20%, Sanger sequencing failed to detect any of the mutations. In comparison, the Boxin NGS assay was able to detect LADRM with theoretical frequencies ranged from 5% to 20%.

Comparison of Boxin NGS with Sanger sequencing using blood samples

Reverse transcriptase (RT) and protease (PR) sequences were genotyped by Sanger sequencing and Boxin NGS. 60/78 samples were identified to have NRTI or NNRTI associated DRMs, and no protease inhibitor (PI) associated DRM was detected. In total, Sanger sequencing identified 109 mutations, and Boxin NGS identified 142 NRTI or NNRTI associated DRMs. Variants detected by Boxin NGS alone were listed in bold. Boxin NGS detected additional 23/33 LADRM with frequencies below 20%. 10/33 mutations had frequencies higher than 20%; 2/10 mutations had frequencies greater than 50% (*table.2*). Same amino acid variants caused by different changes of nucleic acids detected by Boxin NGS were underlined.

Comparison of Boxin NGS with Vela NGS using blood samples

15 blood samples were sequenced by all three methods. RT/PR sequences underwent sanger sequencing and deep-sequencing, DRMs associated with PI, NRTI, and NNRTI drugs were identified. Corresponding variants frequencies were given by both Vela and Boxin NGS assays.

Overall, a total number of 20 mutations were identified by either or both NGS platform(s). Sanger sequencing only detected 5/20 mutations with frequencies higher than 20%. Boxin NGS detected 18/20 variants and Vela NGS detected 13/20 variants. 2 LADRM with frequencies below 10% were detected by Vela NGS alone. 7 LADRM with frequencies below 10% were reported by Boxin NGS alone (*table 3*).

Boxin NGS and Vela NGS detected 11 common mutations, and the numerical values were analyzed statistically. A Bland-Altman plot was graphed to demonstrate the agreement of the numerical data between the two methods (*Fig.2*). The lower limit of agreement was -0.252, the upper limit of agreement was 0.213, the bias was -0.0194. All points distributed close to the line of bias, and no point exceeded the limit of agreement.

Four blood samples were collected from a patient at different periods between 2016 and 2017 (*table 4*). In the second quarter of 2016, no variant was detected or reported by neither sequencing method. However, at the end of year 2016, Boxin NGS detected a minority variant M184V with a frequency of

4.92%, while Sanger sequencing and Vela NGS failed to detect this low-frequency variant. Boxin NGS detected this LADRM 3 months prior to other sequencing platforms. In the 1st quarter of 2017, Sanger sequencing reported a non-specific result as M184MV, and the Vela assay reported the mutant M184V with a frequency of 53.59%. Half a year later, the variant frequency reached 99%.

Discussion

In this study, we evaluated the Boxin NGS assay, and compared the performance of the Boxin NGS assay with the conventional Sanger sequencing method and the FDA-approved Vela NGS assay.

To evaluate the accuracy, reproducibility, and reliability of the Boxin NGS assay, 10 plasmid samples were made with different variant frequencies. Samples were made and examined in triplicates, and the experiment was repeated by two examiners who worked independently. The theoretical values were graphed against the experimental results as illustrated in figure 1A. The slope of the linear regression line was 0.8948, close to 1; the observed values also distributed close to predicted values, suggesting they were highly concordant. The coefficient of determination R^2 was 0.9932, suggesting variables had an excellent correlation. Reproducibility was defined as the variation of the same measurement under different conditions²⁵. The reproducibility of the Boxin NGS assay was visualized by comparing the two sets of measurements carried out by two examiners using the Bland-Altman plot^{25,26}. All data points distributed closely to the line of bias, suggesting the assay had an excellent reproducibility (*Fig. 1B*). The degree of inaccuracy between repeated measurements could be expressed by the coefficient of variation. The average coefficient of variation was calculated to be 7%, suggests that the results had an excellent reliability and accuracy²⁵.

60/78 blood samples underwent Sanger sequencing and Boxin NGS were identified to have NRTI or NNRTI associated DRMs, and no protease inhibitor (PI) associated DRM was detected. Compared to Sanger sequencing, Boxin NGS detected additional 33 DRMs, 23/33 mutations with frequencies below 20%. Sanger sequencing also failed in detecting two mutations with frequencies above 50%. Sanger sequencing also gave non-specific results such as Y188YL, M184MV. This could happen when peaks share a similar intensity. Boxin NGS also reported the frequency of the same amino acid variant caused by different changes of nucleic acids as shown in table 2. Compared to Sanger sequencing, Boxin NGS was able to specific and numerical results, and the platform was more sensitive in detecting LADRM.

15 random blood samples collected from different AIDS patients underwent all three sequencing methods. Compared to Sanger sequencing, Vela NGS and Boxin NGS were sensitive enough to detect and quantify LADRM with frequencies below 20%. The Bland-Altman plot was widely used to illustrate the agreement between two instruments or two methods^{9,25-28}, and was applied to determine the degree of agreement of Boxin NGS and Vela NGS (*Fig. 2*). The average value of the variant frequencies reported by both platforms were plotted against the difference between the same measurements. The scattering points distributed close to the line of bias without exceeding the lines of limit of agreement,

implying that the two methods were highly concordant^{25,29-31}. 5 additional LADRM variants with frequencies at or below 10% were detected by Boxin NGS, suggesting that Boxin NGS platform performed better at detecting LADRM variants. As shown in table 4, Boxin NGS detected a low-frequency M184V variant with a frequency of 4.92%, and other methods failed to detect this variant. The result was later confirmed by Sanger sequencing and Vela NGS, three months after the variant was first reported by Boxin NGS. The mutation kept developing and became the dominant mutation of the HIV-1 quasi-species; the frequency reached to 99% in 9 months. Early detection of LADRM variants can significantly reduce the possibility of virological failure, and potentially benefit the treatment outcome. This case highlighted the reliability and sensitivity of Boxin NGS in detecting LADRM variants. Boxin NGS method was also more cost-efficient. The cost per sample for Sanger sequencing ranged from £430 up to £1050³². Vela NGS costed \$400 for each sample⁵, while Boxin NGS only costed \$40 per sample, which was 90% less than Vela NGS.

In conclusion, the new cost-efficient, robust Boxin NGS appeared to be more sensitive in reporting genomic variants, especially for LADRM variants, without compromising the accuracy. The ability of early identifying and predicting DRMs of Boxin NGS allowed it to have more potentials in improving ART outcomes and preventing treatment failures. Although the HIV-1 LADRM variant was confirmed to have correlations with virologic failure¹⁰, more cohort studies should be conducted to clarify the threshold for switching treatment and how to respond to the presence of LADRM variants.

Competing Interests

YB and JM are the shareholders of Liangxin Biotechnology Development LTD. (Beijing, China). The remaining authors declare that they have no competing interests.

Funding

The study was funded by Liangxin Biotechnology Development LTD. (Beijing, China).

JC reports receiving Zhejiang Provincial Natural Science Foundation (grant no. QY20H190002).

Acknowledgments

We thank Beijing Ditan Hospital Capital Medical University for providing Vela NGS results. We appreciate their hard work and their diligence to patient care.

Author Contributions

YB, MZ, JC, ZY, YW, and WN designed the study. YB, MZ, JC, ZY, YW, WN, LT, WD, and XW performed molecular and genotyping experiments. YB, MZ, LT, JM, WD, and XW provided the logistics in implementing the Boxin NGS assay. CZ, MZ, and LT analyzed the data. CZ wrote and drafted the manuscript. All authors read and validated the manuscript.

References:

1. Thuy Le *et al.* Low-Abundance HIV Drug-Resistant Viral Variants in Treatment-Experienced Persons Correlate with Historical Antiretroviral Use. *PLoS ONE* **4**, e6079 (2009).
2. Bin Lin *et al.* HIV drug resistance in HIV positive individuals under antiretroviral treatment in Shandong Province, China. *PLoS ONE* **12**, e0181997 (2017).
3. Jessica M. Conway & Alan S. Perelson. Early HIV infection predictions: role of viral replication errors. *SIAM J Appl Math* **78**, 1863–1890 (2018).
4. Hezhao Ji *et al.* Bioinformatic data processing pipelines in support of next-generation sequencing-based HIV drug resistance testing: the Winnipeg Consensus. *J. Int. AIDS Soc.* **21**, e25193 (2018).
5. Hezhao Ji *et al.* Are We Ready for NGS HIV Drug Resistance Testing? The Second “Winnipeg Consensus” Symposium. *Viruses* **12**, 586 (2020).
6. Maria Casadellà & Roger Paredes. Deep sequencing for HIV-1 clinical management. *Virus Res.* **239**, 69–81 (2017).
7. Ying Zhou, Jing Lu, Jing Wang, Hongjing Yan, & Haiyang Hu. Prevalence of HIV Antiretroviral Drug Resistance and Its Impacts on HIV-1 Virological Failures in Jiangsu, China: A Cross-Sectional Study. *BioMed Res. Int.* **2016**, 1752437 (2016).
8. David Bonsall *et al.* A Comprehensive Genomics Solution for HIV Surveillance and Clinical Monitoring in Low-Income Settings. *J. Clin. Microbiol.* **58**, e00382-20 (2020).
9. Emma R. Lee *et al.* Performance comparison of next generation sequencing analysis pipelines for HIV-1 drug resistance testing. *Sci. Rep.* **10**, 1634 (2020).
10. Nicholas Silver *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**, (2018).
11. Xianwu Pang *et al.* HIV drug resistance and HIV transmission risk factors among newly diagnosed individuals in Southwest China. *BMC Infect. Dis.* **21**, 160 (2021).
12. Pleuni S. Pennings. HIV drug resistance: problems and perspectives. **5**, s1e5 (2013).
13. Andrea Arias, Pablo López, Raphael Sánchez, Yasuhiro Yamamura, & Vanessa Rivera-Amill. Sanger and Next Generation Sequencing Approaches to Evaluate HIV-1 Virus in Blood Compartments. *Int. J. Environ. Res. Public Health* **15**, 1697 (2018).

14. Keylie M. Gibson, Margaret C. Steiner, Seble Kassaye, Frank Maldarelli, & Keith A. Crandall. A 28-Year History of HIV-1 Drug Resistance and Transmission in Washington, DC. *Front Microbiol* **10**, 369 (2019).
15. Jiayi Chen *et al.* HIV-1 Drug Resistance, Distribution of Subtypes, and Drug Resistance-Associated Mutations in Virologic Failure Individuals in Chengdu, Southwest China, 2014-2016. *BioMed Res. Int.* **2020**, 5894124 (2020).
16. Philip L. Tzou *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**, e00105-18 (2018).
17. Neil T. Parkin *et al.* Multi-Laboratory Comparison of Next-Generation to Sanger-Based Sequencing for HIV-1 Drug Resistance Genotyping. *Viruses* **12**, 694 (2020).
18. Matthew M. Cousins, San-San Ou, Maria J. Wawer, Supriya Munshaw, & Andrew D. Redd. Comparison of a High-Resolution Melting Assay to Next-Generation Sequencing for Analysis of HIV Diversity. *J. Clin. Microbiol.* **50**, 3054–3059 (2012).
19. Julie Yamaguchi, Ana Olivo, Oliver Laeyendecker, Kenn Forberg, & Michael G. Berg. Universal Target Capture of HIV Sequences From NGS Libraries. *Front Microbiol* **9**, 2150 (2018).
20. Stephanie Raymond *et al.* HIV-1 genotypic resistance testing using the Vela automated next-generation sequencing platform. *J. Antimicrob. Chemother.* **73**, 1152–1157 (2018).
21. Stéphanie Raymond *et al.* Performance evaluation of the Vela Dx Sentosa next-generation sequencing system for HIV-1 DNA genotypic resistance. *J. Clin. Virol.* **122**, 104229 (2020).
22. Minjie Chu *et al.* HIV-1 CRF01_AE strain is associated with faster HIV/AIDS progression in Jiangsu Province, China. *Sci. Rep.* **7**, 1570 (2017).
23. Zesong Sun *et al.* Natural polymorphisms in HIV-1 CRF01_AE strain and profile of acquired drug resistance mutations in a long-term combination treatment cohort in northeastern China. *BMC Infect. Dis.* **20**, 178 (2020).
24. Lin Li *et al.* CRF07_BC Strain Dominates the HIV-1 Epidemic in Injection Drug Users in Liangshan Prefecture of Sichuan, China. *AIDS Res. Hum. RETROVIRUSES* **31**, 5 (2015).
25. Karina V. Bunting, Richard P. Steeds, Luke T. Slater, & Dipak Kotecha. A Practical Guide to Assess the Reproducibility of Echocardiographic Measurements. *JASE* **32**, 1505–1515 (2019).
26. J. W. Bartlett & C. Frost. Reliability, repeatability and reproducibility: analysis of measurement errors in continuous variables. *Ultrasound Obstet. Gynecol.* **2008**, 466–475 (2008).

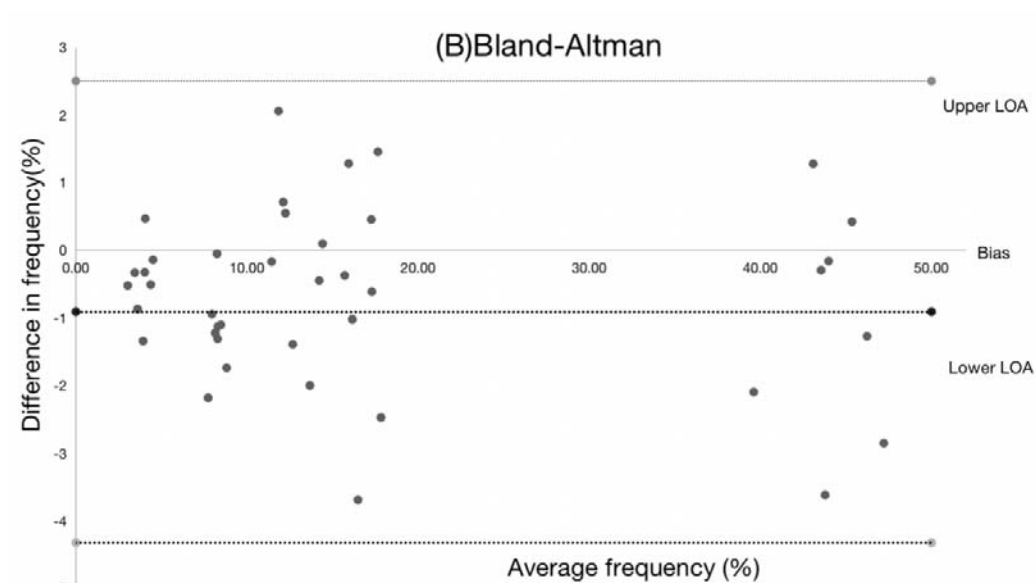
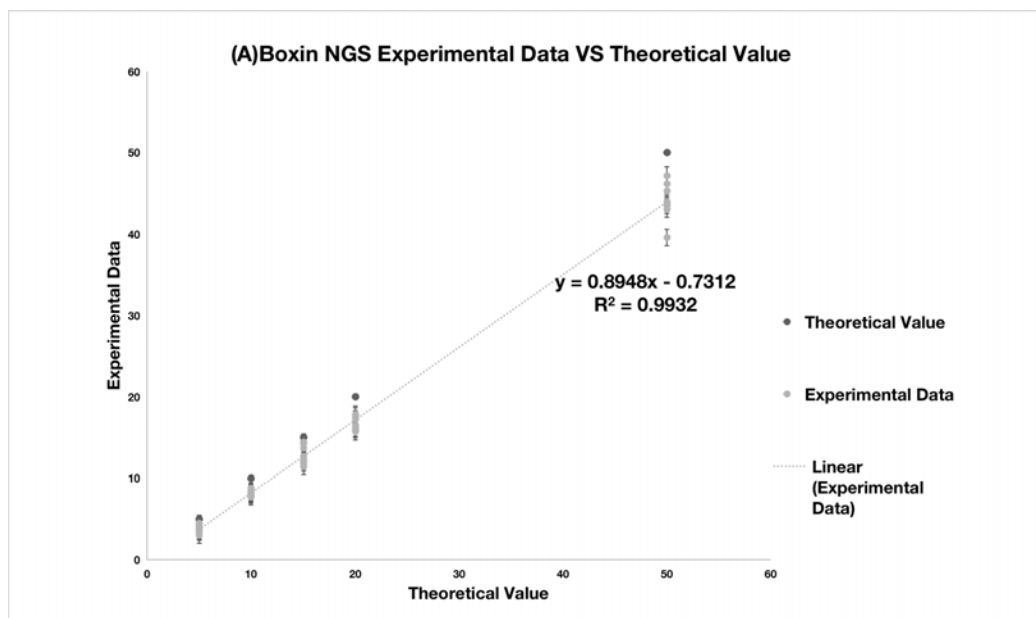
27. Fusako Fujimura, Kazutaka Kamiya, Kazuko Fujiwara, & Kimiya Shimizu. Repeatability and Reproducibility of Measurements Using a NT-530P Noncontact Tono/Pachymeter and Correlation of Central Corneal Thickness with Intraocular Pressure. *BioMed Res. Int.* **2013**, 370592 (2013).
28. Oke Gerke. Reporting Standards for a Bland–Altman Agreement Analysis: A Review of Methodological Reviews. *Diagn. Basel* **10**, 334 (2020).
29. Nurettin Özgür Doğan. Bland-Altman analysis: A paradigm to understand correlation and agreement. *Turk. J. Emerg. Med.* **18**, 139–141 (2018).
30. Catey Bunce. Correlation, Agreement, and Bland–Altman Analysis: Statistical Analysis of Method Comparison Studies. *Am. J. Ophthalmol.* **148**, 4–6.
31. David Preiss & Joseph Fisher. A Measure of Confidence in Bland–Altman Analysis for the Interchangeability of Two Methods of Measurement. *J. Clin. Monit. Comput.* **22**, 257–259 (2008).
32. Next Generation DNA Sequencing: A Review of the Cost Effectiveness and Guidelines [Internet]. *Ott. Can. Agency Drugs Technol. Health* (2014).

Appendix

Sample	YB002(μ L)	YB003(μ L)	YB012(μ L)	Theoretical frequency(%)
1	95	5	-	5
2	90	10	-	10
3	85	15	-	15
4	80	20	-	20
5	50	50	-	50
6	95	-	5	5
7	90	-	10	10
8	85	-	15	15

9	80	-	20	20
10	50	-	50	50

Table 1. Constructing samples with known variant frequencies. YB002: wild type; YB003: including single amino acid variant, K65R; YB012: including seven amino acid variants, L100I, K101E, V106M, Y181F, Y188L, G190E, and M230L



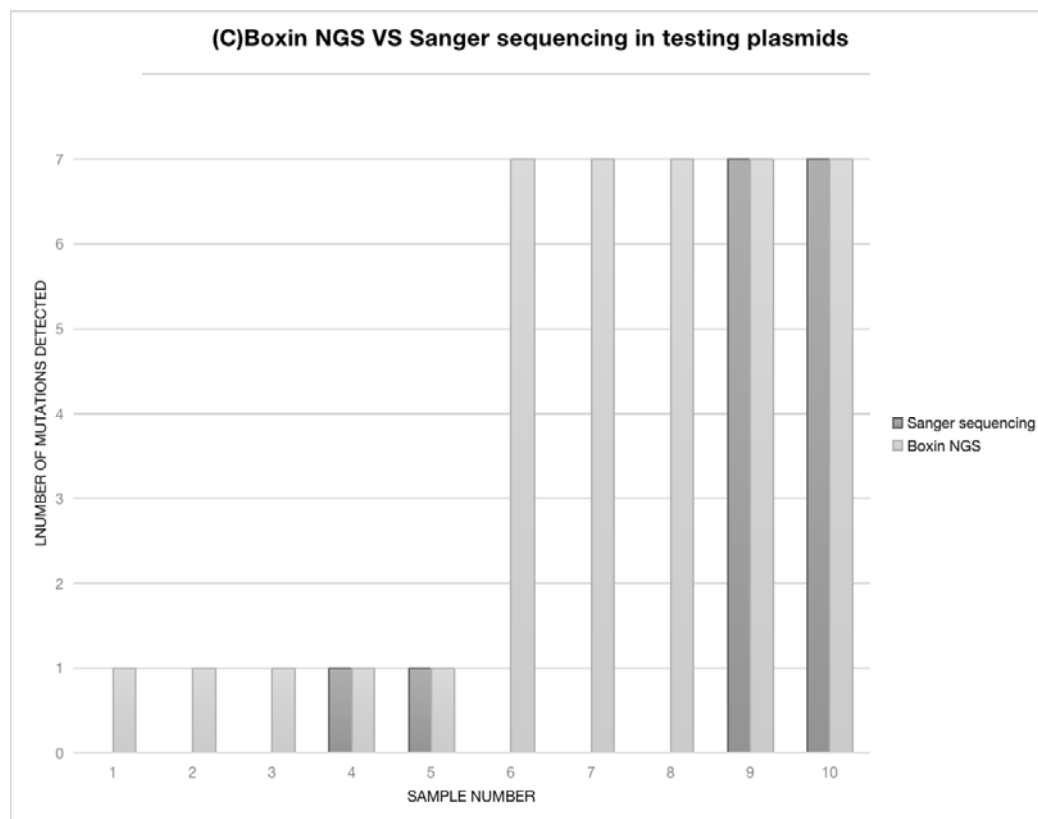


Figure 1. (A) Experimental data VS theoretical predictions of the Boxin NGS assay. The theoretical values were represented on the x-axis (theoretical mutation frequencies 5%, 10%, 15%, 20%, and 50%). The experimental results of the mutation frequencies with standard errors were labelled on the y-axis. (B) The Bland-Altman plot. Comparison of the Boxin NGS assay carried out by two examiners. The y-axis represented the frequency difference for a particular variant detected by two examiners. The x-axis represented the average frequency for a particular variant detected by two examiners. The x-axis represented the average frequency for a particular variant detected by two examiners.

Sample	Sanger		Boxin NGS	
	NRTI	NNRTI	NRTI	NNRTI
R-RT-BX2025			Y181C(9.4%); H221Y(4.62%)	
R-RT-BX2031			K101E(8.64%)	
R-RT-BX2046			I47V(3.45%)	
R-RT-BX2049		K103KN	K103N(43.15%);G190E(3.46%)	

R-RT-BX2053			K101E(7.94%)	
R-RT-BX2056	M184V	K103N, H221Y, M230L, L234I	M184V(93.6%); <u>M184V(6.1%)</u> ;	K103N(98.24%);H221Y(98.3%);M230L(99.55%); L234I(98.47%); T215Y(1.53%)
R-RT-BX2057		K103KN	D30N(2.95%) ;K103N(78.51%); G190E(10.05%)	
R-RT-BX2058			K103N(91.49%)	
R-RT-BX2059	M184V	H221Y, M230L, K238N	K103N(21.94%) ; V108I(33.64%) ; M184V(99.65%)	H221Y(55.79%);M230L(99.85%);K238N(99.42%)
R-RT-BX2060		Y181C	Y181C(91.75%); <u>Y181C(3.57%)</u> ; H221Y(33.75%)	
R-RT-BX2061			M184V(4.92%)	
R-RT-BX2063			M184V(16.71%)	
R-RT-BX2064	M184V, T215Y	K103N, Y181C	M41L(23.51%) ; D67G(27.81%) ; K101H(22.61%) ; M184V(99.7%); G190A(26.33%) ; T215Y(99.95%)	K103N(62.32%); Y181C(99.45%);
R-RT-BX2066		K103N, V179E	K103N(82.63%); V179E(86.18%); G190A(3.57%)	
R-RT-BX2067			T215A(5.91%)	
R-RT-BX2070			V106A(53.99%) ; Y181C(5.95%)	
R-RT-BX2072		V179D	V179D(94.65%); <u>V179D(5.09%)</u> ; M184V(4.92%)	
R-RT-BX2073			K103N(19.08%) ; P225H(14.48%)	
R-RT-BX2074			M41L(13.8%) ; V106I(22.43%)	
R-RT-BX2075		K103N	K103N(87.5%); <u>K103N(9.8%)</u> ; Y181C(4.75%); K219E(4.04%)	
R-RT-BX2078	M184V		K70R(4.96%) ; K101E(9.11%) ; M184V(48.22%)	
R-RT-BX2084		K103N	D67G(2.91%) ; K103N(100%)	

Table 2. Discordant results between Sanger sequencing and the Boxin NGS assay in terms of NRTI and NNTRI associated DRMs. No PI DRM was identified by either of the two methods. Variants which were detected by Boxin NGS assay only were highlighted in bold. Same amino acid variants detected by Boxin NGS but caused by different mutated nucleic acids were underlined.

Mutation Frequency	Boxin	Vela
1%-5%	5	2
5%-10%	3	2
10%-20%	2	0
>20%	8	9
Total	18	13

Table 3. Number of mutations detected by Boxin and Vela NGS platform.

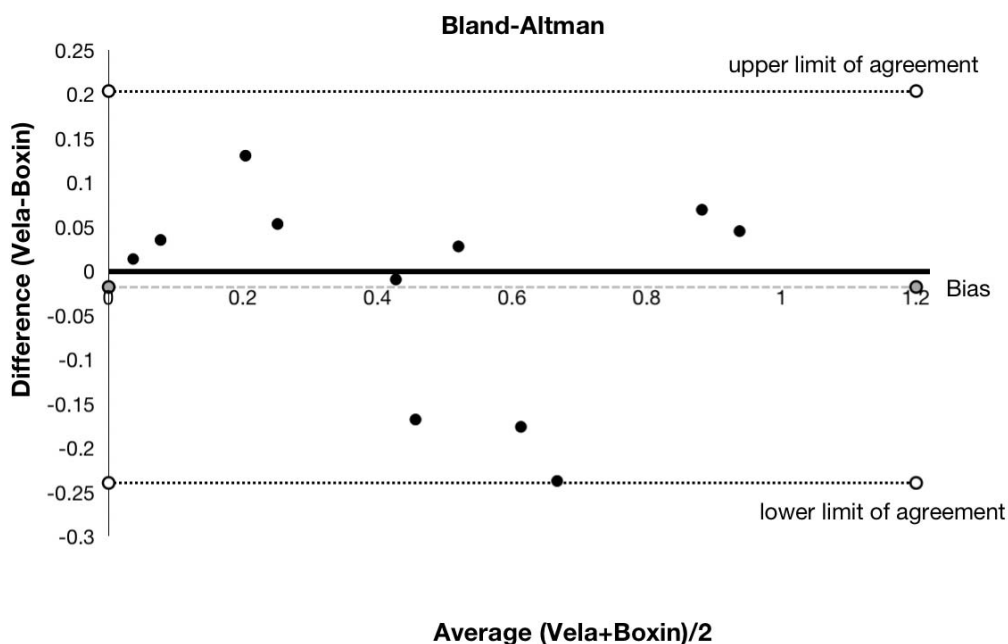
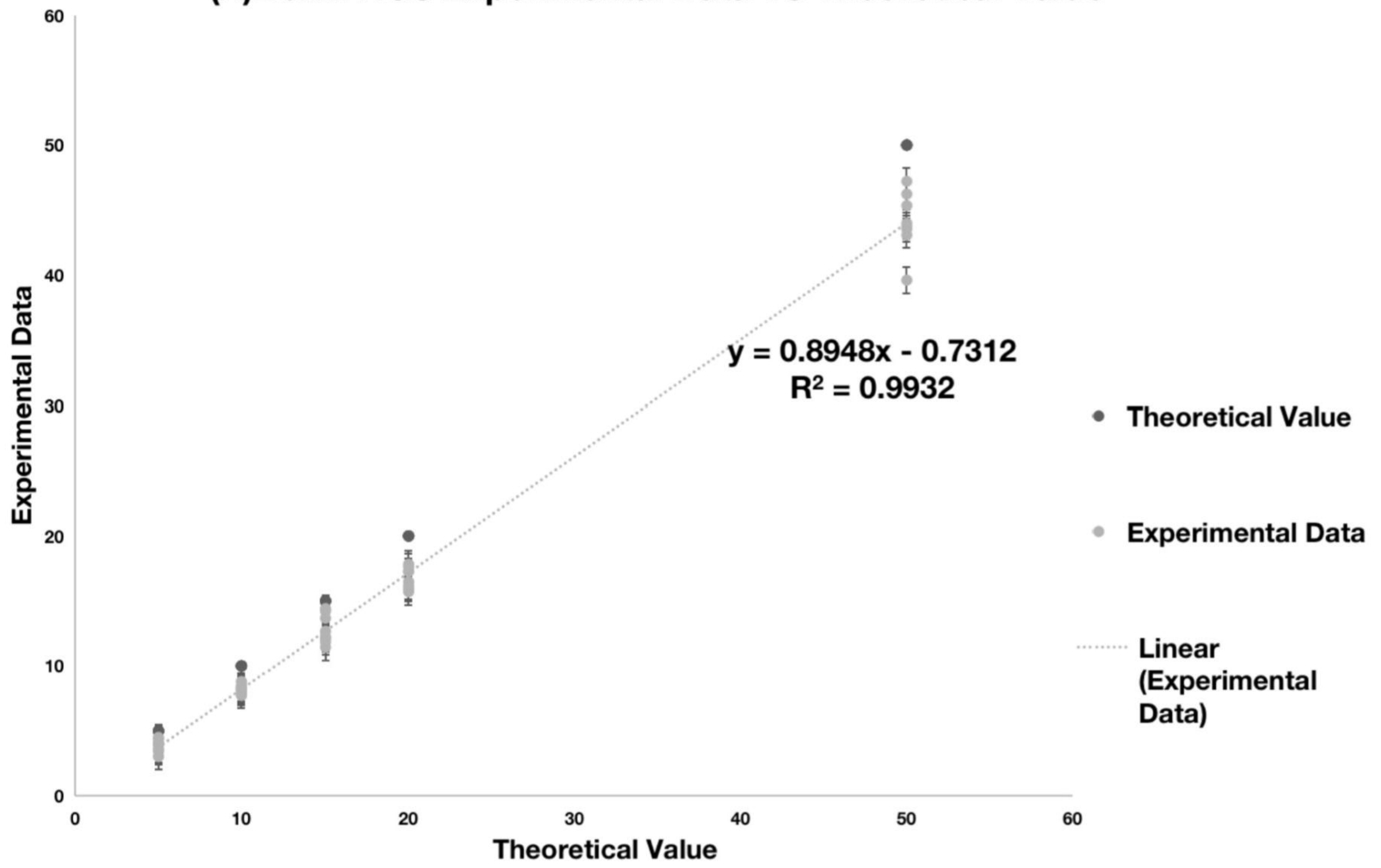


Figure 2. The Bland-Altman plot. Comparison of Boxin and Vela NGS methods in quantifying drug-resistance associated mutations. The y-axis represented the variant frequency difference detected by Boxin and Vela NGS. The x-axis represented the average variant frequency detected by Boxin and Vela NGS.

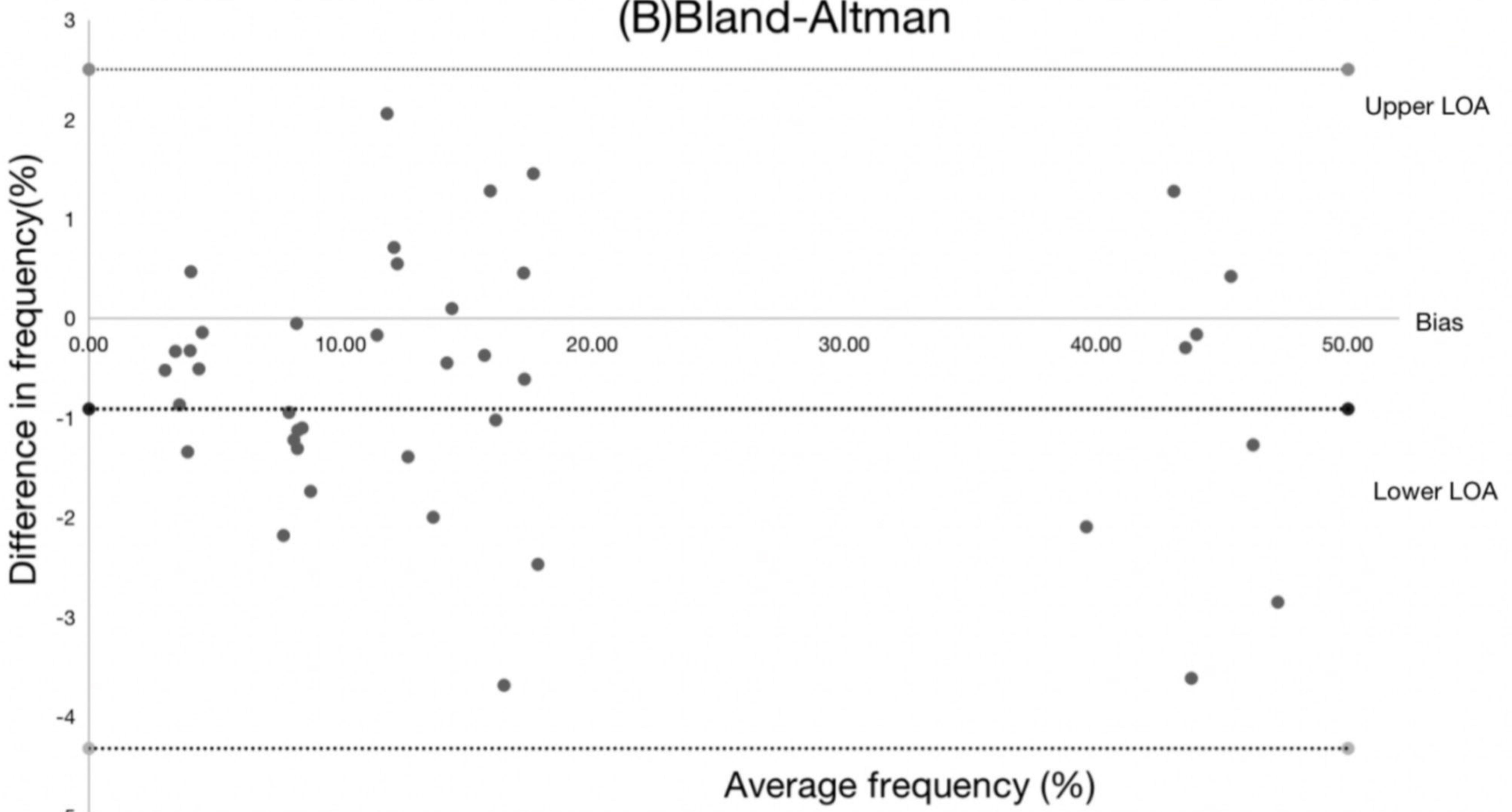
HIV-1 infected case 3	SS	Vela NGS	Boxin NGS
2016 2nd Quarter	None	None	None
2016 4th Quarter	None	None	M184V(4.92%)
2017 1st Quarter	M184MV	M184V(53.39%)	M184V(50.58%)
2017 3rd Quarter	M184V	M184V(99.01%)	M184V(99.45%)

Table 4. Mutations detected in blood samples collected from a single patient in different periods between 2016 and 2017.

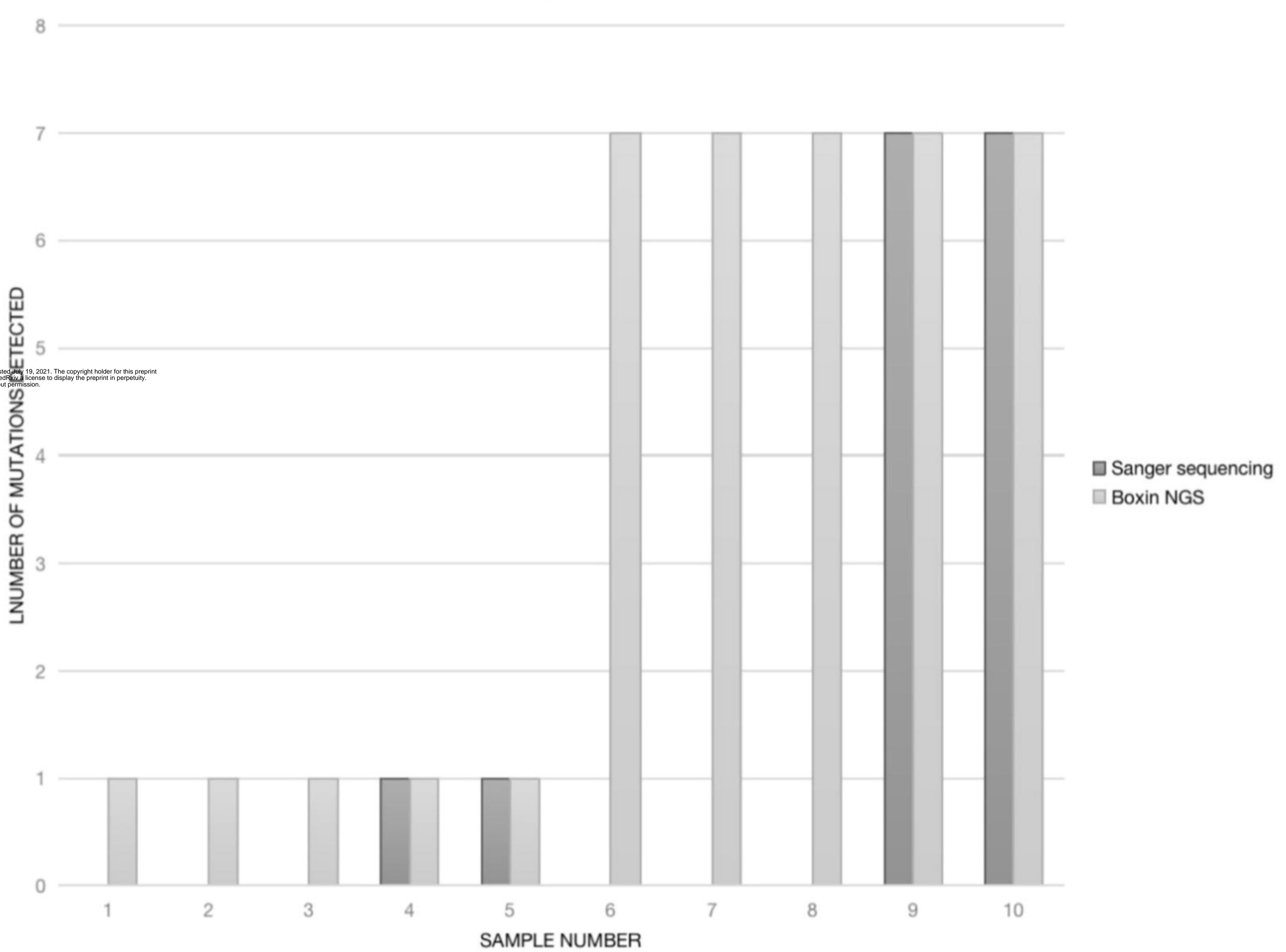
(A) Boxin NGS Experimental Data VS Theoretical Value



(B) Bland-Altman



(C) Boxin NGS VS Sanger sequencing in testing plasmids



Bland-Altman

