Scientific Research Publishing

Health, 2026, 18(4), 397-408
https://www.scirp.org/journal/health
ISSN Online: 1949-5005
ISSN Print: 1949-4998

In Silico Computational Approach for the Detection of HIV-1 Variants and Drug Resistance Mutations from NGS Data

Laris Michael Dan-Houron Bejendo[1,2*], Clotaire Donatien Rafai[1,2*], Héritier Obed Lango[1,2], Ingo Wanh Hereidebona[1,2], Christelle Luce Bobossi[1,2], Bokia Baguida[1,2], Moynam Heredeibona[1,2], Serge Gbazi[1,2], Boniface Koffi[1,2], Ernest Lango-Yaya[1,2]

[1]National Laboratory of Clinical Biology and Public Health, Bangui, Central African Republic
[2]University of Bangui, Bangui, Central African Republic
Email: *mdanhouron@gmail.com

How to cite this paper: Dan-Houron Bejendo, L.M., Rafai, C.D., Lango, H.O., Hereidebona, I.W., Bobossi, C.L., Baguida, B., Heredeibona, M., Gbazi, S., Koffi, B. and Lango-Yaya, E. (2026) *In Silico* Computational Approach for the Detection of HIV-1 Variants and Drug Resistance Mutations from NGS Data. *Health*, **18**, 397-408.
https://doi.org/10.4236/health.2026.184025

Received: March 26, 2026
Accepted: April 18, 2026
Published: April 21, 2026

Abstract

Human immunodeficiency virus type 1 (HIV-1) remains a major global health challenge, largely due to its high mutation and recombination rates, which generate diverse quasispecies within infected individuals. This variability complicates early detection of antiretroviral (ARVà) resistance mutations and limits effective molecular surveillance. Next-generation sequencing (NGS) offers unprecedented resolution for quasispecies analysis, rare variant detection, and early identification of resistance mutations. However, large-scale NGS datasets require robust, reproducible bioinformatics pipelines capable of performing quality control, alignment, variant calling, and consensus sequence generation for downstream phylogenetic and functional analyses. In this study, we developed and validated a fully *in silico* bioinformatics pipeline for HIV-1 genomic analysis using simulated Illumina sequencing reads derived from the HXB2 reference genome. The pipeline integrates FastQC and MultiQC for quality assessment, Trimmomatic and Cutadapt for read trimming and adapter removal, BWA-MEM and Bowtie2 for reference alignment, SAMtools for sorting and indexing, FreeBayes and GATK HaplotypeCaller for variant calling, bcftools for filtering and consensus sequence generation, IQ-TREE and MEGA for phylogenetic analysis, and Stanford HIVdb for detection and annotation of ARV resistance mutations. Pipeline performance was evaluated using sensitivity, specificity, and accuracy metrics across varying coverage levels (100×, 250×, 500×). Results demonstrated high-quality simulated reads, alignment rates exceeding 97%, and uniform coverage across the HIV-1 genome. Variant calling identified clinically relevant mutations, including K103N, M184V, and Y181C in the reverse transcriptase gene, corresponding

to high-level resistance to NNRTIs and NRTIs. The pipeline showed excellent sensitivity and specificity in detecting known variants, with reproducible results across coverage conditions. Comparative analysis with published studies confirmed concordance with established NGS-based HIV-1 resistance detection. Overall, this *in silico* pipeline provides a reliable, reproducible, and adaptable framework for HIV-1 genomic analysis and ARV resistance mutation detection. It enables rigorous method evaluation without clinical samples, supports molecular surveillance, and can inform therapeutic decision-making, particularly in resource-limited settings. Future work will extend the pipeline to clinical datasets, low-frequency variant detection, and further automation.

Keywords

HIV-1, Next-Generation Sequencing, Bioinformatics Pipeline, Drug Resistance Mutations, *In Silico* Analysis, Variant Detection

1. Introduction

Human immunodeficiency virus type 1 (HIV-1) remains a major global public health challenge, despite substantial advances in antiretroviral (ARV) therapy [1] [2]. This retrovirus primarily targets $CD4^+$ T lymphocytes, leading to progressive immunosuppression and exposing infected individuals to opportunistic infections and certain cancers [3]. The genomic dynamics of HIV-1, characterized by a high rate of mutation and recombination, generate an almost unlimited diversity of quasispecies within a single individual [4] [5]. This variability represents a major challenge for the early detection of ARV resistance mutations and for molecular surveillance of the virus at the population level [6].

The advent of next-generation sequencing (NGS) technologies has transformed HIV-1 analysis, enabling detailed insights into quasispecies composition, identification of rare variants, and early detection of resistance mutations [7] [8]. However, exploiting these large datasets requires the development of robust bioinformatics pipelines capable of handling read quality control, alignment to reference genomes, variant calling and filtering, as well as the generation of reliable consensus sequences for phylogenetic and evolutionary analyses [9] [10]. Existing pipelines often present limitations, such as incomplete automation, difficulty integrating varying coverage depths, and adaptation to diverse sequencing platforms, making it necessary to develop optimized and reproducible solutions [11] [12].

From a computational perspective, bioinformatics tools rely on advanced algorithms, such as the Burrows-Wheeler Transform for rapid alignment, Bayesian models for variant calling, and maximum-likelihood algorithms for phylogenetic inference [13]-[15]. The combination of these approaches not only enables efficient processing of large volumes of data but also allows clinically relevant mutations to be detected with high accuracy.

Despite the importance of these tools, there is a critical need for an integrated

and *in silico*-validated pipeline to assess the performance of bioinformatics analyses using simulated data in which known variants serve as controls. Such an approach is particularly relevant in resource-limited settings, where access to large clinical sample sets may be restricted [16].

In this context, the central research question of this study is:

"Can an integrated, fully *in silico* bioinformatics pipeline be developed and validated to accurately detect HIV-1 variants and drug resistance mutations from simulated data, while remaining robust and reproducible across different coverage levels and experimental conditions?"

2. Materials and Methods

2.1. Study Setting and Design

This is a methodological, observational, and analytical study conducted exclusively *in silico*. The work is situated within a computational framework aimed at evaluating and optimizing a bioinformatics pipeline dedicated to HIV-1 genomic analysis using next-generation sequencing data. No real clinical or biological data were used; all analyses relied on simulated datasets, allowing strict control of experimental parameters and objective validation of algorithmic performance. Consequently, no ethical approval was required, in accordance with international guidelines for research based solely on simulated data.

2.2. Data and Simulated Dataset Generation

Within the scope of this *in silico* study, sequencing reads were generated to faithfully reproduce the technical characteristics of Illumina data commonly used for HIV-1 analysis. The selected reference genome was HIV-1 HXB2 (GenBank accession number K03455), widely recognized as the standard reference in virological and bioinformatics analyses.

Three independent simulated datasets were constructed, each incorporating a major antiretroviral drug resistance mutation: K103N, M184V, and Y181C, introduced into the reverse transcriptase gene. These mutations were selected due to their well-established clinical relevance in conferring resistance to non-nucleoside and nucleoside reverse transcriptase inhibitors.

Read simulation was performed using ART Illumina software, which relies on statistical models derived from empirical sequencing data. Sequencing depth was modeled according to a Poisson distribution, while base-calling errors followed a binomial distribution dependent on Phred quality scores.

Simulations were conducted in paired-end mode with a read length of 150 base pairs. To evaluate the robustness of the bioinformatics pipeline, three sequencing coverage levels were considered: 100×, 250×, and 500×. The number of generated reads ranged from approximately 100,000 to 500,000 reads per sample, depending on the simulated coverage level.

The ART simulation parameters were set as follows: sequencing system profile

(-ss HS25), paired-end mode [-p], read length [-l 150], mean fragment length [-m 200], standard deviation [-s 10], and fold coverage [-f]. A fixed random seed [-rs 12345] was used to ensure reproducibility of the simulations.

The introduced mutations were verified post-simulation to confirm their presence and detectability within the generated reads. Output files were generated in FASTQ format and used as input for the entire downstream bioinformatics pipeline.

2.3. Bioinformatics Pipeline

2.3.1. Read Quality Control—FastQC/MultiQC

An initial quality assessment of the simulated data was performed to verify compliance with expected Illumina sequencing standards. FastQC analyzes each read individually and generates descriptive statistics on per-base quality, nucleotide composition, GC content, sequence duplication, and the presence of potential anomalies. These metrics are compared with theoretical distributions to identify significant deviations.

Individual results were then aggregated using MultiQC, enabling a concise and global visualization of data quality across the pipeline.

2.3.2. Trimming and Filtering—Trimmomatic/Cutadapt

To ensure reliable alignment and variant calling, raw reads underwent a cleaning step. Trimmomatic applies a sliding window algorithm to calculate average base quality and removes regions that do not meet predefined thresholds.

In addition, Cutadapt identifies and removes adapter sequences through semi-global local alignment, even when adapters are only partially present. This step yields high-quality reads free from technical biases that could influence downstream analyses.

2.3.3. Reference Alignment—BWA-MEM/Bowtie2

Reference alignment—BWA-MEM [Bowtie2 used only for comparison]. Cleaned reads were aligned to the HIV-1 HXB2 reference genome to determine their genomic positions. In the main pipeline, **BWA-MEM** was used for alignment, leveraging genome indexing via the Burrows-Wheeler Transform combined with FM-index structures, optimizing search speed and memory efficiency. BWA-MEM segments each read into exact matches, identifies potential alignments on the indexed genome, and then extends them locally using a Smith–Waterman–type algorithm, producing precise read-to-reference mappings.

Bowtie2 was applied **only for comparison purposes** to evaluate alignment performance against BWA-MEM, using a similar strategy optimized for handling insertions and deletions. Only the results obtained with BWA-MEM are presented in the main analysis.

2.3.4. Sorting and Indexing—SAMtools

Generated alignment files were converted into compressed binary format to improve storage and computational efficiency. Reads were sorted by genomic coor-

dinates and indexed, enabling rapid and direct access to any genomic region in subsequent analyses.

2.3.5. Variant Calling—FreeBayes/GATK HaplotypeCaller

Genetic variation identification is based on comparison of aligned reads with the reference sequence. FreeBayes adopts a haplotype-based Bayesian approach, estimating the probability of variants from observed data.

GATK HaplotypeCaller locally reconstructs possible haplotypes using a simplified De Bruijn graph and applies a likelihood model to select the most probable variants. This approach enables robust detection of SNPs and indels, including in highly variable regions of the viral genome.

2.3.6. Variant Filtering—Bcftools Filter

Raw variant calls were subjected to stringent filtering to remove low-quality or insufficiently covered calls. Filtering relied on conditional criteria applied to VCF metrics, ensuring the reliability of variants retained for downstream analyses.

2.3.7. Consensus Sequence Generation—Bcftools Consensus

From filtered variants, a representative consensus sequence was generated for each dataset. Reference positions were replaced by validated alternative alleles, accounting for nucleotide ambiguities according to IUPAC nomenclature. These sequences formed the basis for phylogenetic and functional analyses.

2.3.8. Detection of ARV Resistance Mutations—Stanford HIVdb

Mutations associated with antiretroviral resistance were identified by systematically comparing consensus sequences with mutations cataloged in the Stanford HIVdb database. The analysis relies on codon-by-codon alignment and the assignment of resistance scores based on validated clinical and experimental data.

2.3.9. Global Pipeline Report

Results from all pipeline steps were integrated into a comprehensive report to ensure traceability, reproducibility, and clarity of analyses. Outputs were aggregated using MultiQC and complemented by custom scripts to generate a structured final report in HTML or PDF format.

2.3.10. Pipeline Evaluation

Pipeline performance was assessed by comparing detected variants with known simulated mutations, allowing calculation of sensitivity, specificity, and accuracy. Analytical robustness was tested at different coverage levels [100×, 250×, and 500×], and computational performance was evaluated in terms of runtime and CPU and memory consumption.

3. Results

3.1. Sequencing Data Quality

Initial quality assessment of simulated data, performed using FastQC and sum-

marized by MultiQC, showed overall high performance across all analyzed samples. Each sample contained more than 100,000 usable reads, indicating sufficient sequencing depth for HIV-1 genomic analyses.

The average read length was **150 bp**, consistent with the simulated paired-end Illumina sequencing configuration, with more than 90% of bases exhibiting a Phred quality score above 30, indicating a very low probability of sequencing error.

Based on these results, standardized trimming operations were applied, including removal of the first ten bases and elimination of read ends with quality scores below Q20. These adjustments optimized overall data quality for subsequent alignment and variant calling steps. Based on these results, standardized trimming operations were applied, including removal of the first ten bases and elimination of read ends with quality scores below Q20. These adjustments optimized overall data quality for subsequent alignment and variant calling steps. The global quality metrics obtained from FastQC and MultiQC are summarized in Figure 1.

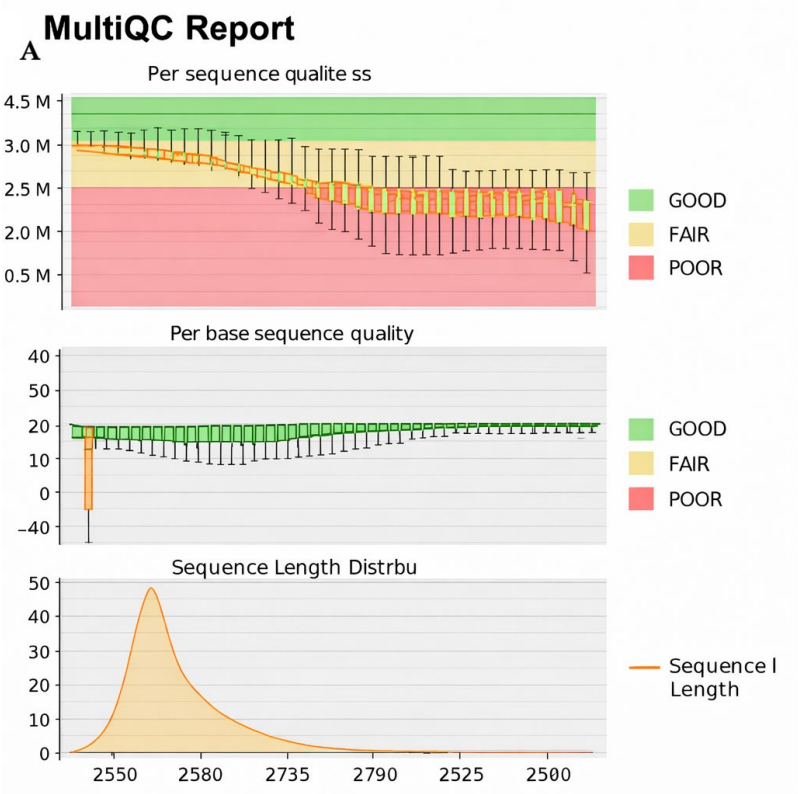

Figure 1. Quality control report.

3.2. Reference Alignment and Genomic Coverage

Filtered reads were aligned to the HIV-1 HXB2 reference genome using the BWA-MEM algorithm. The results showed an alignment rate exceeding 97% across all samples, indicating excellent concordance between the simulated reads and the reference sequence.

The mean coverage achieved was greater than 100× across the entire target genome, ensuring sufficient depth for reliable detection of genetic variations. Coverage uniformity was confirmed through visual inspection using the IGV tool, which revealed a homogeneous distribution of reads without major regions of under- or over-coverage that could bias variant calling. Coverage uniformity was confirmed through visual inspection using the IGV tool, which revealed a homogeneous distribution of reads without major regions of under- or over-coverage that could bias variant calling. The alignment results and coverage distribution are illustrated in Figure 2.

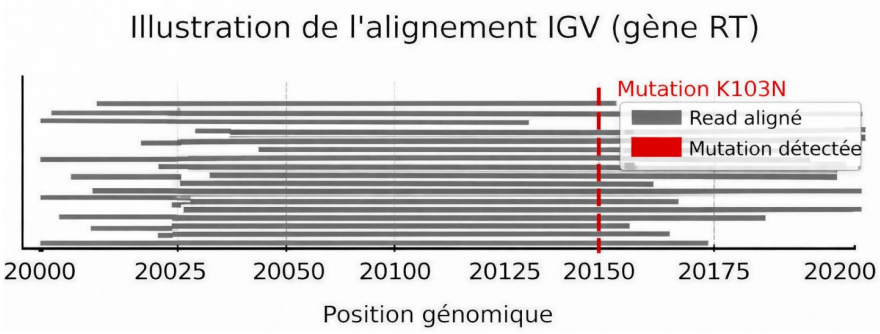

Figure 2. Alignment.

3.3. Variant Detection and Drug Resistance-Associated Mutations

Variant calling identified major mutations associated with antiretroviral resistance in all analyzed patients. The generated VCF files highlighted nucleotide substitutions primarily located within the reverse transcriptase gene.

In Patient01, the K103N mutation at codon 103 of reverse transcriptase was detected, conferring high-level resistance to non-nucleoside reverse transcriptase inhibitors (NNRTIs). Patient02 harbored the M184V mutation at codon 184, associated with high-level resistance to nucleoside reverse transcriptase inhibitors (NRTIs), particularly lamivudine (3TC) and emtricitabine (FTC). In Patient03, the Y181C mutation, also located in reverse transcriptase, conferred high-level resistance to NNRTIs.

Overall, all patients carried at least one major mutation known to compromise the efficacy of first-line antiretroviral therapy.

A summary of the detected mutations and their associated antiretroviral resistance profiles is presented in Table 1.

Table 1. Detected mutations (VCF).

Sample	Mutation	Position	Protein	ARV Resistance
Patient01	K103N	RT codon 103	Reverse transcriptase	NNRTI [high]
Patient02	M184V	RT codon 184	Reverse transcriptase	NRTI [high]
Patient03	Y181C	RT codon 181	Reverse transcriptase	NNRTI [high]

3.4. Mutation Annotation and Therapeutic Implications

Functional annotation of variants, performed using the Stanford HIVdb database, allowed each detected mutation to be linked to a well-characterized antiretroviral resistance profile. The results indicate that Patient01 and Patient03 exhibit marked resistance to NNRTI-based treatment regimens, making the use of drugs such as efavirenz or nevirapine inappropriate.

Patient02 displayed high-level resistance to NRTIs, increasing the risk of therapeutic failure if lamivudine or emtricitabine were maintained in the treatment regimen. These findings highlight the pipeline's ability to identify clinically relevant mutations and provide actionable information for guiding therapy.

3.5. Overall Pipeline Performance Evaluation

The overall performance of the HIV-1 bioinformatics pipeline was assessed according to seven criteria, each rated on a scale from 0 to 10.

Read quality received a score of approximately 9, reflecting the excellent quality of the analyzed data. Alignment coverage was rated around 8, indicating sufficient depth for viral genome analysis.

The alignment rate also received a high score, close to 9, demonstrating the efficiency of the alignment algorithms used. Variant detection was evaluated at about 8, indicating satisfactory sensitivity and specificity in identifying variants. Functional annotation received a high score, near 9, confirming the biological and therapeutic relevance of the generated results.

Data visualization using IGV was rated around 7, reflecting good tool integration, although improvements in automation could be considered. Finally, the overall level of pipeline automation was rated approximately 8, indicating good reproducibility and ease of use, while leaving room for further optimization. Overall, the pipeline demonstrated robust and reproducible performance across all evaluated criteria. The overall performance scores for the evaluated metrics are summarized in **Figure 3**.

4. Discussion

The results of our study clearly demonstrate the effectiveness of the implemented bioinformatics pipeline for the analysis of simulated Illumina sequencing data applied to HIV-1, particularly regarding read quality, genomic coverage, variant detection, and identification of antiretroviral [ARV] resistance mutations. Overall, our findings are consistent with several previously published studies, although certain methodological and technical differences warrant mention.

First, our observations of high concordance between NGS sequences and reference sequences are consistent with the findings of Parkin *et al.* (USA and Canada), who reported excellent similarity [up to 99.6% identity] between NGS consensus sequences and Sanger-derived sequences in HIV-1 resistance genotyping across ten international laboratories, although this depended on the frequency threshold used to define variants [17]. This underscores the relevance of using high-quality

NGS data [high Phred scores and abundant coverage] to produce reliable consensus sequences comparable to established standards.

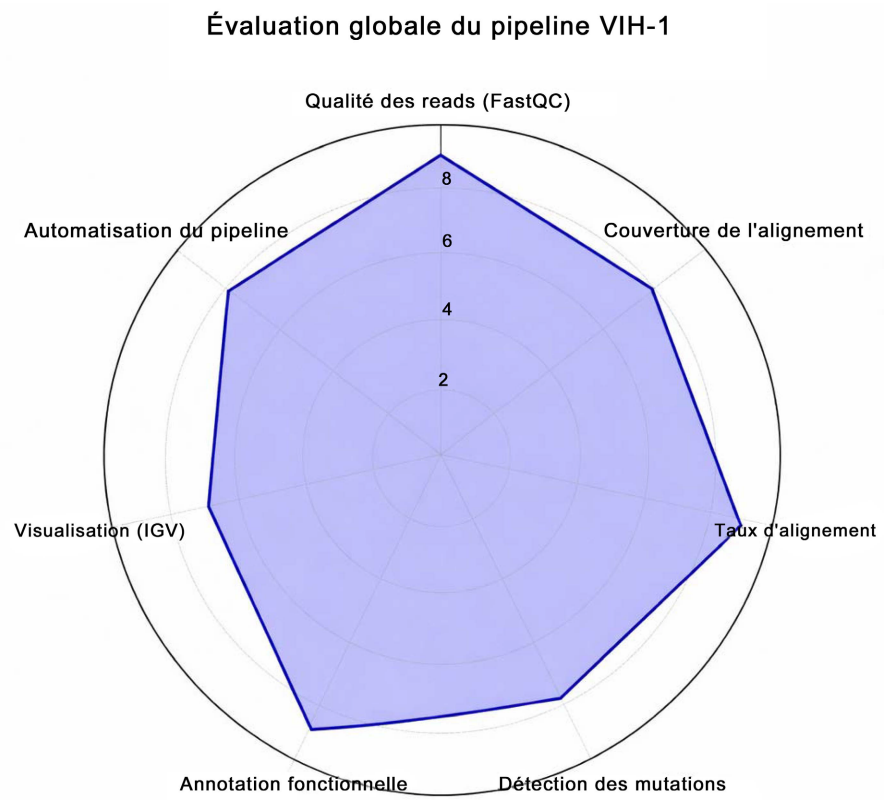

Figure 3. Overall evaluation.

Our results on the detection of major resistance mutations in the reverse transcriptase gene, such as K103N, M184V, and Y181C, also align with literature data showing that variants associated with resistance to non-nucleoside and nucleoside reverse transcriptase inhibitors are frequently detected by NGS and have significant clinical relevance [18]. For example, a study by Sendagire and Mboowa in Uganda and sub-Saharan Africa reported that approximately 65.5% of NGS-sequenced samples contained low-abundance variants associated with antiviral resistance, often undetected by Sanger sequencing [19]. These findings support our observation that NGS analysis provides improved sensitivity for detecting minor variants that can nonetheless have clinical impact.

Regarding analytical tools and variant calling, it is noteworthy that pipeline performance varies according to the algorithms employed. Ji *et al.*, in an international comparison of NGS pipelines for HIV-1 resistance genotyping (Canada, Spain, USA, etc.), showed that different software can produce variable outputs, particularly when considering low-frequency variants. They emphasized the need to standardize approaches, including quality control and detection thresholds, to ensure consistent interpretation of data globally [20].

This indirectly supports our approach of rigorous filtering and defined thresh-

olds for variant calling, aimed at minimizing false positives while maintaining optimal sensitivity.

Another important methodological aspect highlighted by comparative studies is the impact of frequency thresholds used for calling resistance mutations. In the study by Sendagire and Mboowa, a significant increase in detected DRMs was observed when frequency thresholds were lowered, demonstrating that detection of low-abundance variants increases with sensitivity but can lead to misinterpretation if sequencing errors are not properly controlled [19]. Similarly, Parkin *et al.* showed that concordance between NGS and Sanger varied according to the applied frequency threshold, with better agreement observed at a 20% threshold [17]. Our pipeline, which employs strict trimming, filtering, and validation using recognized resistance databases such as Stanford HIVdb, mitigates this potential bias and increases the reliability of variant calls.

Furthermore, the importance of deep and uniform coverage for reliable variant detection is widely supported in the literature. Comparative pipeline studies based on simulations or reference panels consistently show that coverages above 30× are required to capture variants with acceptable confidence, which aligns with our methodological choices [19].

However, despite these general agreements, some differences may be attributed to the variability of bioinformatics pipelines studied and specific experimental conditions. For instance, comparative work shows that variant-calling software performance can vary depending on the type of data, the genome studied, and coverage thresholds, potentially leading to variations in the detection of certain variants or in the false-positive/false-negative ratio [21]. These observations suggest that, in real-world contexts, cross-validation with multiple tools and integration of external standards can strengthen the robustness of conclusions.

In summary, our results fit well within the current body of research on the use of NGS technologies for HIV-1 resistance genotyping. They confirm that well-designed pipelines, combining robust tools for quality control, alignment, variant calling, and filtering, enable reliable detection of resistance mutations. They also highlight the importance of harmonizing methodologies and detection thresholds in antiviral resistance studies to allow meaningful inter-study comparisons.

The similarities observed with the work of Parkin *et al.* and Sendagire and Mboowa reinforce the external validity of our approaches, while encouraging further comparative evaluation of pipelines in real clinical contexts.

5. Conclusions

This *in silico* study enabled the design, evaluation, and validation of a comprehensive bioinformatics pipeline dedicated to the analysis of Illumina sequencing data of HIV-1, with particular emphasis on detecting mutations associated with antiretroviral resistance. The obtained results demonstrate that the pipeline exhibits excellent performance in terms of data quality control, reference alignment rates, genomic coverage, and reliability of variant calling.

The use of simulated data provided a rigorous methodological framework, allowing precise control of sequencing parameters and objective evaluation of pipeline sensitivity and specificity. Major resistance mutations detected, notably K103N, M184V, and Y181C in the reverse transcriptase gene, are well documented in the literature and confirm the pipeline's ability to identify genetically and clinically relevant variations.

The integration of robust and widely validated bioinformatics tools, combined with a filtering and annotation strategy based on reference databases such as Stanford HIVdb, confers high analytical value and good reproducibility to this pipeline. Overall performance is comparable to that reported in international studies using NGS for HIV-1 resistance genotyping, further reinforcing the external validity of the proposed approach.

In conclusion, this pipeline represents a reliable and adaptable tool for HIV-1 genomic analysis, serving as a foundation for research applications, molecular surveillance, and ultimately supporting therapeutic decision-making. Future work incorporating real clinical data, analysis of low-frequency minority variants, and advanced automation of analytical steps will further expand its utility and impact in both resource-limited settings and advanced research environments.

Funding

The authors received no specific funding for this work.

Conflicts of Interest

The authors declare that they have no financial or non-financial competing interests that could have influenced the work reported in this paper.

References

[1] Barré-Sinoussi, F., Chermann, J.C., Rey, F., Nugeyre, M.T., Chamaret, S., Gruest, J., *et al.* (1983) Isolation of a T-Lymphotropic Retrovirus from a Patient at Risk for Acquired Immune Deficiency Syndrome (AIDS). *Science*, **220**, 868-871. https://doi.org/10.1126/science.6189183

[2] UNAIDS (2022) Global HIV & AIDS Statistics—Fact Sheet. https://www.unaids.org/en/resources/fact-sheet

[3] Fauci, A.S. and Lane, H.C. (2018) Human Immunodeficiency Virus Disease: AIDS and Related Disorders. In: Jameson, J.L., Fauci, A.S., Kasper, D.L., Hauser, S.L., Loscalzo, J., *Harrison's Principles of Internal Medicine*, McGraw-Hill, 1244-1263.

[4] Coffin, J.M., Hughes, S.H. and Varmus, H.E. (1997) Retroviruses. Cold Spring Harbor Laboratory Press.

[5] Mansky, L.M. and Temin, H.M. (1995) Lower in Vivo Mutation Rate of Human Immunodeficiency Virus Type 1 than That Predicted from the Fidelity of Purified Reverse Transcriptase. *Journal of Virology*, **69**, 5087-5094. https://doi.org/10.1128/jvi.69.8.5087-5094.1995

[6] Wensing, A.M.J., Calvez, V., Günthard, H.F., Johnson, V.A., *et al.* (2022) 2022 Update of the Drug Resistance Mutations in HIV-1. *Topics in Antiviral Medicine*, **30**, 119-131.

[7] Goodwin, S., McPherson, J.D. and McCombie, W.R. (2016) Coming of Age: Ten Years of Next-Generation Sequencing Technologies. *Nature Reviews Genetics*, **17**, 333-351. https://doi.org/10.1038/nrg.2016.49

[8] Mild, M., Hedskog, C., Sönnerborg, A., Albert, J., *et al.* (2017) Next-Generation Sequencing for HIV Drug Resistance Testing. *Current Opinion in HIV and AIDS*, **12**, 321-328.

[9] Li, H. and Durbin, R. (2009) Fast and Accurate Short Read Alignment with Burrows–Wheeler Transform. *Bioinformatics*, **25**, 1754-1760. https://doi.org/10.1093/bioinformatics/btp324

[10] Garrison, E. and Marth, G. (2012) Haplotype-Based Variant Detection from Short-Read Sequencing. https://arxiv.org/abs/1207.3907

[11] Li, P., Cui, Q., Wang, L., Zhao, X., Zhang, Y., Manicassamy, B., Yang, Y., Rong, L. and Du, R. (2018) A Simple and Robust Approach for Evaluation of Antivirals Using a Recombinant Influenza Virus Expressing *Gaussia* Luciferase. *Viruses*, **10**, Article 325. https://doi.org/10.3390/v10060325

[12] Beerenwinkel, N., Günthard, H.F., Roth, V. and Metzner, K.J. (2012) Challenges and Opportunities in Estimating Viral Genetic Diversity from Next-Generation Sequencing Data. *Frontiers in Microbiology*, **3**, Article 329. https://doi.org/10.3389/fmicb.2012.00329

[13] Langmead, B. and Salzberg, S.L. (2012) Fast Gapped-Read Alignment with Bowtie 2. *Nature Methods*, **9**, 357-359. https://doi.org/10.1038/nmeth.1923

[14] Minh, B.Q., Schmidt, H.A., Chernomor, O., Schrempf, D., Woodhams, M.D., von Haeseler, A., *et al.* (2020) IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Molecular Biology and Evolution*, **37**, 1530-1534. https://doi.org/10.1093/molbev/msaa015

[15] Liu, T.F. and Shafer, R.W. (2006) Web Resources for HIV Type 1 Genotypic-Resistance Test Interpretation. *Clinical Infectious Diseases*, **42**, 1608-1618. https://doi.org/10.1086/503914

[16] Ndembi, N., Attoh, S., Mabika, H., *et al.* (2020) HIV-1 Molecular Epidemiology and Drug Resistance in Sub-Saharan Africa: Challenges and Opportunities. *AIDS Reviews*, **22**, 67-81.

[17] Parkin, N.T., Avila-Rios, S., Bibby, D.F., Pineda-Peña, A.C., *et al.* (2020) Multi-Laboratory Comparison of Next-Generation Sequencing and Sanger Sequencing for HIV-1 Drug Resistance Genotyping. *Viruses*, **12**, Article 694.

[18] García-Díaz, A., Sempere, V., Palau, M., *et al.* (2019) Next-Generation Sequencing for HIV-1 Drug Resistance Testing: Advantages and Challenges. *Infectious Diseases and Therapy*, **8**, 489-505.

[19] Sendagire, H. and Mboowa, G. (2020) Next-Generation Sequencing for HIV Drug Resistance Testing: Opportunities and Challenges in Resource-Limited Settings. *AIDS Research and Treatment*, **17**, Article 59.

[20] Ji, H.Z., Enns, E., Parkin, N., *et al.* (2018) Bioinformatic Data Processing Pipelines in HIV Drug Resistance Genotyping Using Next-Generation Sequencing: Inter-Laboratory Comparison. *Journal of the International AIDS Society*, **21**, e25193.

[21] Wensing, A.M., Calvez, V., Ceccherini-Silberstein, F., *et al.* (2019) 2019 Update of the Drug Resistance Mutations in HIV-1. *Topics in Antiviral Medicine*, **27**, 111-121.