Generated: 2026-01-19 01:26:49

Statistical Methodology Guide

This report uses four advanced statistical methods combined with drug-specific clinical thresholds to ensure accurate, clinically reliable results. This page explains each method in plain English, why we use it, and how to interpret the results.

1. Confidence Intervals (95% CI)

What it is:

A confidence interval gives us a range where the TRUE resistance prevalence likely falls. Think of it as a 'margin of error' around our measurement.

Why we use it:

A single percentage (e.g., '2.5% resistance') doesn't tell us how CONFIDENT we are. The CI tells us: 'We're 95% sure the true value is between X% and Y%'.

How to interpret it:

Example: NVP shows 2.5% resistance with CI [1.8% - 3.4%]
- NARROW interval (1.8-3.4): High confidence, good data
- If CI crosses threshold: Result is UNCERTAIN
- If entire CI is above threshold: Strong evidence of resistance

2. Multiple Testing Correction (FDR)

What it is:

When we test 27 drugs, we're doing 27 separate tests. This increases the chance of false positives. FDR correction adjusts p-values to control the overall error rate.

Why we use it:

Without correction: Testing 27 drugs at 5% error rate each gives ~75% chance of at least ONE false positive! With FDR, we control the OVERALL error rate at 5%.

How to interpret it:

Example: Drug shows p=0.03 (raw) but p=0.08 (FDR-adjusted)
- Raw test says 'significant' but FDR says 'not significant after correction'
- ALWAYS use the FDR-adjusted p-value for clinical decisions
- This report tested 7 drugs simultaneously

3. Statistical Power Analysis

What it is:

Power tells us: IF resistance exists at the threshold level, what's the probability we'll detect it? It measures whether we have ENOUGH data to find resistance if it's there.

Why we use it:

If we say 'no resistance found,' we need to prove we had enough data. Low power means: 'We might have missed it due to insufficient sequencing depth.'

How to interpret it:

This analysis: Average Power = 0.00 (0%) - LOW - Need more sequencing
- Power >= 0.80: Gold standard, results trustworthy
- Power 0.60-0.79: Acceptable, but borderline
- Power < 0.60: CAUTION - May miss real resistance

Note: Power varies by drug based on its specific threshold

4. Binomial Exact Test (for Rare Variants)

What it is:

When a drug has fewer than 30 resistant reads, the standard Z-test becomes unreliable. The exact test uses precise mathematical formulas instead of approximations.

Why we use it:

For rare variants (e.g., 5 reads out of 1,000), the Z-test can give WRONG p-values. The exact test is mathematically correct for all sample sizes.

How to interpret it:

In the results table, you'll see which test was used:
- 'Exact': For counts < 30 (more accurate for rare variants)
- 'Z-test': For counts >= 30 (equally accurate, faster)

Both test: 'Is prevalence significantly above the drug-specific threshold?'

5. Drug-Specific Clinical Thresholds

What it is:

Different drugs have different 'barriers to resistance.' This pipeline uses drug-specific thresholds based on clinical evidence rather than a one-size-fits-all cutoff.

Why we use it:

Example thresholds:
- NNRTIs (e.g., Efavirenz): 1% (single mutation can cause resistance)
- NRTIs (e.g., 3TC): 20% (M184V highly prevalent but tolerable)
- PIs (e.g., Darunavir): 15% (requires multiple mutations)

Using the correct threshold for each drug improves clinical relevance.

How to interpret it:

In the results:
- Each drug's threshold is shown in the detailed table
- Drug-specific threshold lines appear on the prevalence plot
- Significance is determined relative to EACH drug's threshold

Summary: Ensuring Clinical Accuracy

These five methods work together:
1. CIs quantify CONFIDENCE in each measurement
2. FDR correction prevents FALSE POSITIVES from multiple testing
3. Power analysis validates NEGATIVE results
4. Exact tests ensure ACCURACY for rare variants

5. Drug-specific thresholds provide CLINICAL RELEVANCE

This approach meets publication standards and ensures results are clinically actionable and scientifically defensible.

HIV Drug Resistance Mutation Analysis Report

Generated: 2026-01-19 01:26:49

Executive Summary

Total Reads Processed: 1,600

Usable High-Quality Reads: 6

Analysis Duration: 0m 26s

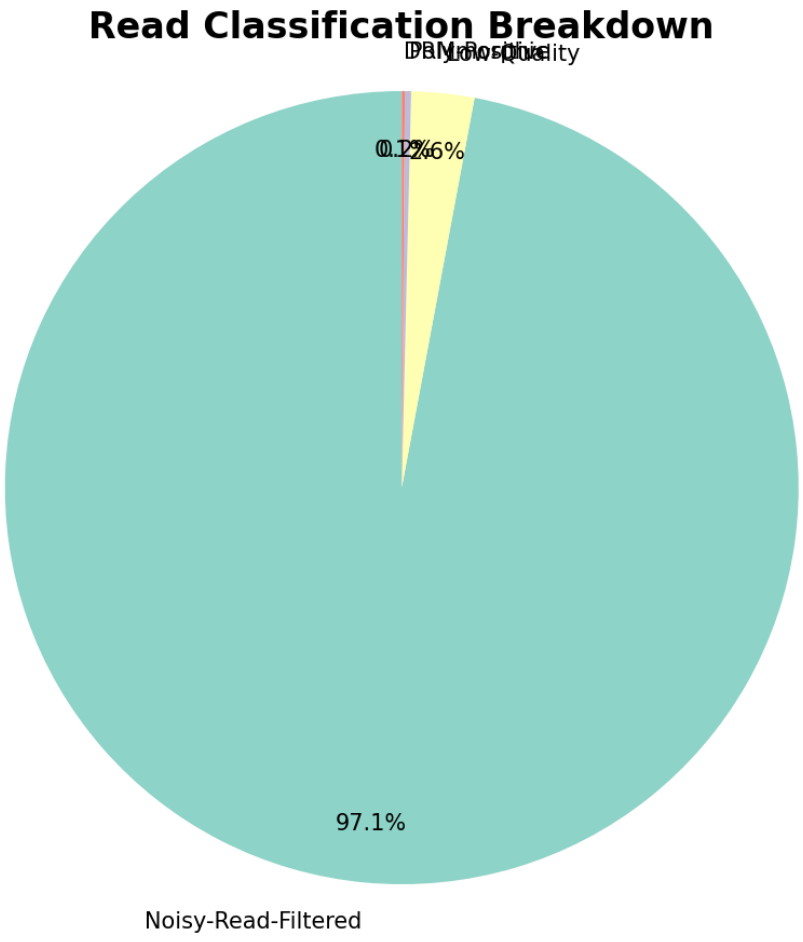
Usable Read Percentage: 0.4%

Average Statistical Power: 0.00 (0%)

Low - May miss low-frequency resistance. Consider more sequencing.

Read Classification:

- Noisy-Read-Filtered: 1,553 (97.1%)
- Low-Quality: 41 (2.6%)
- Polymorphic: 4 (0.2%)
- DRM-Positive: 2 (0.1%)

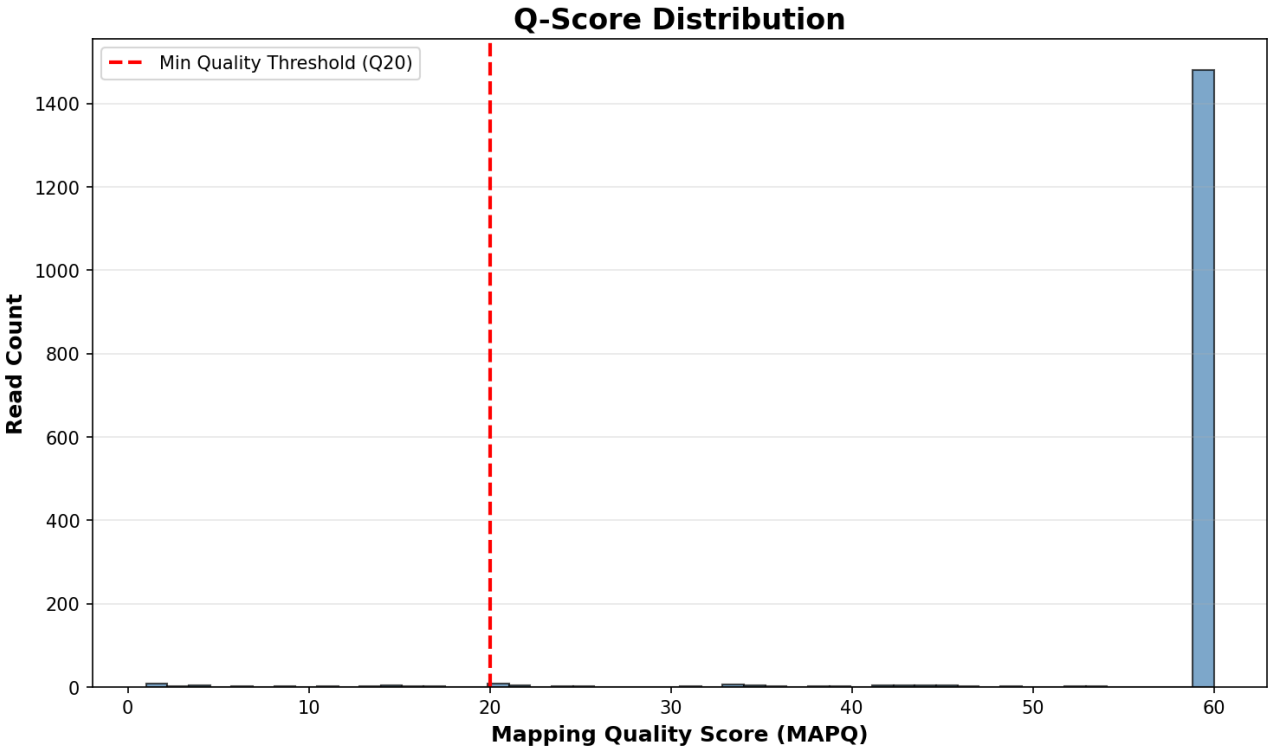


Sequencing Quality Summary

This section shows the quality metrics of the sequencing run, including mapping quality distribution and genome coverage depth. High-quality data ensures reliable mutation calling.

Q-Score Distribution

This plot shows the distribution of mapping quality scores (MAPQ). Higher scores indicate more confident alignments. Reads with MAPQ < 20 are filtered out.

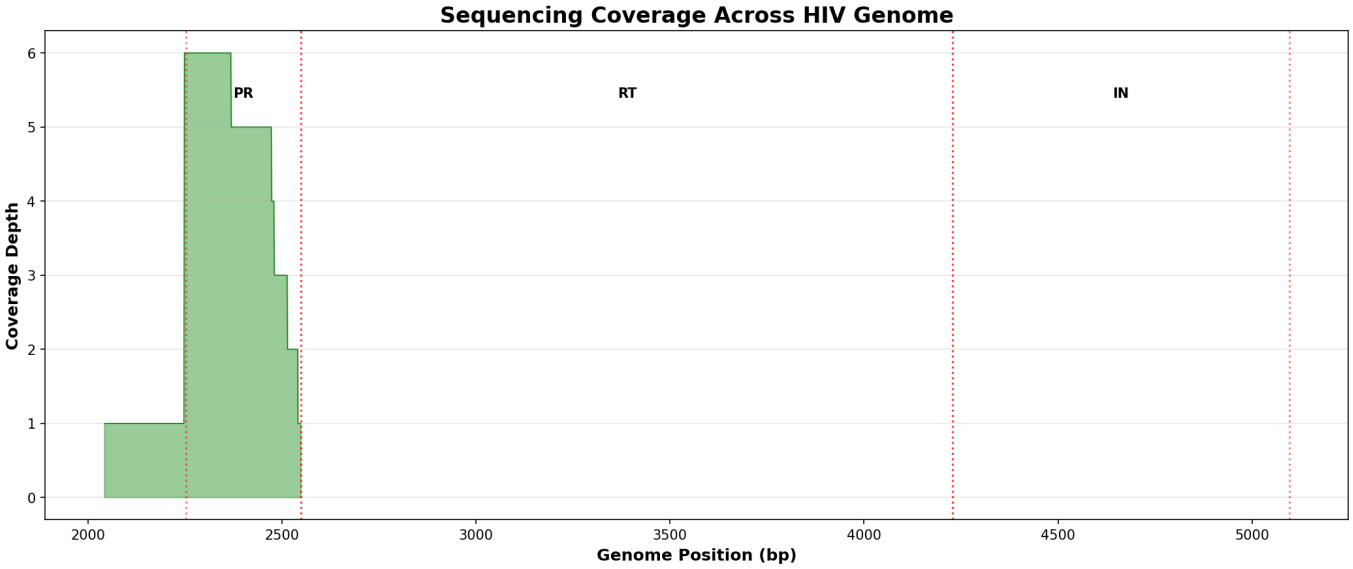


HIV Drug Resistance Mutation Analysis Report

Generated: 2026-01-19 01:26:49

Coverage Depth Across Genome

This plot shows sequencing depth at each position in the HIV genome. Vertical lines mark gene boundaries (PR, RT, IN). Uniform, high coverage ensures reliable mutation detection.



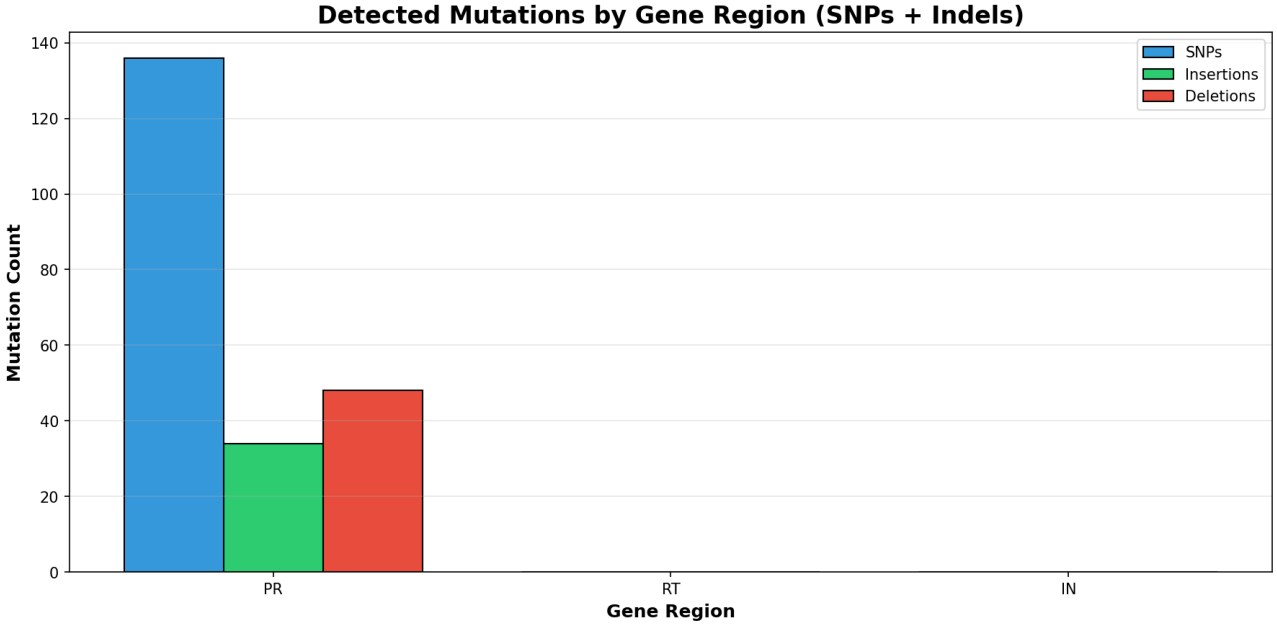
HIV Drug Resistance Mutation Analysis Report

Generated: 2026-01-19 01:26:49

Detected Mutations

Mutation Count by Gene Region (SNPs + Indels)

This chart shows the total number of mutations detected in each gene region, including single nucleotide polymorphisms (SNPs), insertions, and deletions. These include both drug-resistance mutations (DRMs) and natural polymorphisms.



Indel Detection Summary

Insertions and deletions (indels) are important for HIV drug resistance. Notable examples include the T69 insertion complex (NRTI resistance) and various deletions in protease.

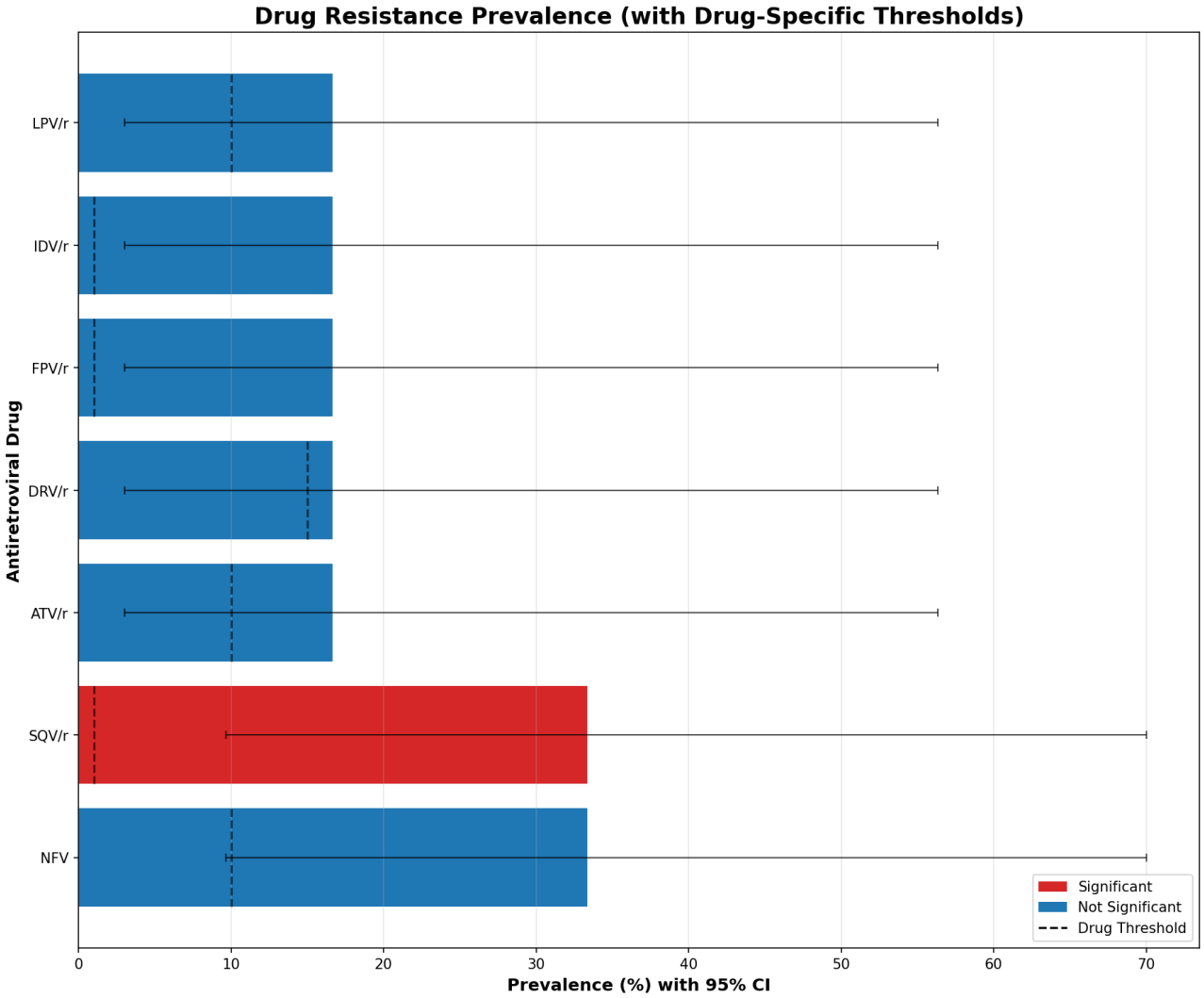
PR: 34 insertions, 48 deletions

HIV Drug Resistance Mutation Analysis Report

Generated: 2026-01-19 01:26:49

Drug Resistance Analysis

The following analysis shows resistance prevalence with 95% confidence intervals. P-values are FDR-corrected for multiple testing. Each drug uses its specific clinical threshold (shown in the table). Drugs marked as 'Significant' have FDR-adjusted p < 0.05 AND prevalence above their drug-specific threshold.



HIV Drug Resistance Mutation Analysis Report

Generated: 2026-01-19 01:26:49

Detailed Statistical Results

Drug	Count	Prevalence	95% CI	Threshold	Power	P (FDR)	Test	Significant
NFV	2	33.33%	[9.7-70.0]	10.0%	0.00	2.00e-01	Exact	NO
SQV/r	2	33.33%	[9.7-70.0]	1.0%	0.00	1.02e-02	Exact	YES
ATV/r	1	16.67%	[3.0-56.4]	10.0%	0.00	5.47e-01	Exact	NO
DRV/r	1	16.67%	[3.0-56.4]	15.0%	0.00	6.23e-01	Exact	NO
FPV/r	1	16.67%	[3.0-56.4]	1.0%	0.00	1.37e-01	Exact	NO
IDV/r	1	16.67%	[3.0-56.4]	1.0%	0.00	1.37e-01	Exact	NO
LPV/r	1	16.67%	[3.0-56.4]	10.0%	0.00	5.47e-01	Exact	NO

Clinical Interpretation

Significant Resistance Detected: 1 Drug(s)

The following drugs showed statistically significant resistance after FDR correction, with prevalence exceeding their drug-specific clinical thresholds. These drugs are likely to have reduced clinical efficacy:

*** SQV/r**

Prevalence: 33.33% [CI: 9.7-70.0%]

Threshold: 1.0% | Power: 0.00 | FDR p: 1.02e-02