Supplementary Material For:

HIV Drug Resistance Testing by High-Multiplex "Wide" Sequencing on the Illumina MiSeq

H Lapointe[a], W Dong[a], GQ Lee[a], DR Bangsberg[b,c,d] , JN Martin[e], AR Mocello [e], Y Boum[d], A Karakas[a], D Kirkby[a], AFY Poon[a,f], PR Harrigan[a,f], CJ Brumme[a]

a. BC Centre for Excellence in HIV/AIDS, Vancouver, Canada
b. Harvard School of Public Health, Boston, MA, USA
c. Massachusetts General Hospital, Boston, MA, USA
d. Mbarara University of Science of Technology, Mbarara, Uganda
e. University of California, San Francisco, CA, USA
f. University of British Columbia, Vancouver, Canada

Corresponding Author:

Chanson Brumme
BC Centre for Excellence in HIV/AIDS
680-1081 Burrard St.
Vancouver, BC Canada, V6Z 1Y6
Phone: (604) 682-2344 ext 63211
FAX: (604) 806-9463
e-mail: cbrumme@cfenet.ubc.ca

Co-Corresponding Author:

Richard Harrigan
Director, Research Labs
BC Centre for Excellence in HIV/AIDS
680-1081 Burrard St.
Vancouver BC Canada V6Z 1Y6
Phone: (604) 806-8281
FAX: (604) 806-8464
e-mail: prharrigan@cfenet.ubc.ca

Table S1: Reverse Transcription and PCR Amplification Primers for MiSeq and Sanger Sequencing of HIV Reverse Transcriptase

Samples	Primer	Step	Direction	HXB2 Location	Sequence (5' to 3')
Canadian-Primary (Sanger)	RT3.1	RT	R	3830-3859	GCTCCTACTATGGGTTCTTTCTCTAACTGG
	5CP1	PCR1	F	1981-2008	GAAGGGCACACAGCCAGAAATTGCAGGG
	2.5	PCR2	F	2011-2039	CCTAGGAAAAAGGGCTGTTGGAAATGTGG
	RT3798R	PCR2	R	3777-3798	CAAACTCCCACTCAGGAATCCA
Canadian-Backup (Sanger)	RT3361R	RT	R	3342-3361	TAAATCTGACTTGCCCAATT
	PRTO5	PCR1	F	2008-2031	GCCCCTAGGAAAAAGGGCTGTTGG
	2.5	PCR2	F	2011-2039	CCTAGGAAAAAGGGCTGTTGGAAATGTGG
	NE1.1	PCR2	R	3303-3323	CTGTATGTCATTGACAGTCCA
UARTO (Sanger)	RTR2	RT	R	3303-3322	TGTATRTCATTGACAGTCCA
	CP2F	PCR1	F	2610-2635	GTTAAACAATGGCCATTGACAGAAGA
	RTF2wd	PCR2	F	2629-2651	CAGAAGARAAAATAAAAGCATTA
	RT3271R	PCR2	R	3252-3271	ACTGTCCATTTRTCAGGATG
MiSeq	ILRT2796F	PCR2	F	2796 - 2815	**TCGTCGGCAGCGTCAGATGTGTATAAGAGACAG**AGAACTCAAGACTTYTGGGA
	ILRT3271R	PCR2	R	3252-3271	**GTCTCGTGGGCTCGGAGATGTGTATAAGAGACAG**ACTGTCCATTTRTCAGGATG

Amplicons generated for Sanger sequencing differed depending on sample source and in the event of PCR failures during initial amplifications. All amplicons were produced by two-step RT-PCR followed by nested PCR amplification (PCR2). Canadian samples were processed by BCCfE clinical staff to generate a contiguous amplicon spanning the entire protease and RT codons 1-400 (Canadian-Primary). A subset of samples had failed a previous amplification attempt and were thus amplified using a 'backup' set of primers to obtain an amplicon spanning protease and RT codons 1-240 (Canadian-Backup). Ugandan samples were all amplified using a single primer set to span RT 35-234 (UARTO). MiSeq amplicons were generated from first-round Sanger products using primers containing Illumina-specific adaptor sequences (bolded bases).

Table S2: Index Sequences for 1152-Fold Multiplex MiSeq Sequencing

Tag	Read	Index Sequence (5’ to 3’)
I521	I1	ACGAGTGC
I522	I1	ACGCTCGA
I523	I1	AGACGCAC
I524	I1	AGCACTGT
I525	I1	ATCAGACA
I526	I1	ATATCGCG
I527	I1	CGTGTCTC
I528	I1	CTCGCGTG
I541	I1	TCTCTATG
I542	I1	TGATACGT
I543	I1	CATAGTAG
I544	I1	CGAGAGAT
I545	I1	ATACGACG
I546	I1	TCACGTAC
I547	I1	CGTCTAGT
I548	I1	TCTACGTA
N501	I1	TAGATCGC
N502	I1	CTCTCTAT
N503	I1	TATCCTCT
N504	I1	AGAGTAGA
N505	I1	GTAAGGAG
N506	I1	ACTGCATA
N507	I1	AAGGAGTA
N508	I1	CTAAGCCT
I731	I2	GTAGTACA
I732	I2	GTAGTCGT
I733	I2	AGTCTACG
I734	I2	TACTCGTA
I735	I2	CGAGAGTA
I736	I2	CGTCTCTA

Tag	Read	Index Sequence (5’ to 3’)
I737	I2	AGCGACGA
I738	I2	GCGTATGT
I739	I2	ACTCGCGT
I740	I2	ATAGTAGT
I741	I2	TGTACAGT
I742	I2	TATAGTCT
I751	I2	ACGACGCT
I752	I2	AGCGTACT
I753	I2	TACTCTAT
I754	I2	GTAGCGTG
I755	I2	GTCTACTG
I756	I2	TCACGTCG
I757	I2	TGTGTGTA
I758	I2	CACGTGTA
I759	I2	GATCTGTA
I760	I2	ACAGCGTA
I761	I2	CTACACTA
I762	I2	GTGATCGA
I771	I2	TAGTGCGA
I772	I2	TCGCTAGA
I773	I2	AGTATAGA
I774	I2	ATACGTCA
I775	I2	TACTCACA
I776	I2	TATACTGT
I777	I2	GATCGCGT
I778	I2	CTGCTAGT
I779	I2	CGTGAGCT
I780	I2	TGTATACT
I781	I2	TCTCGACT
I782	I2	GTAGCACT
N701	I2	TAAGGCGA

Tag	Read	Index Sequence (5' to 3')
N702	I2	CGTACTAG
N703	I2	AGGCAGAA
N704	I2	TCCTGAGC
N705	I2	GGACTCCT
N706	I2	TAGGCATG
N707	I2	CTCTCTAC
N708	I2	CAGAGAGG
N709	I2	GCTACGCT
N710	I2	CGAGGCTG
N711	I2	AAGAGGCA
N712	I2	GTAGAGGA

A short-cycle, indexed PCR enables unique tagging of all amplicons for MiSeq sequencing. The dual-indexing strategy using 24 "forward" (I1) and 48 "reverse" (I2) indices allows barcoding up to 1152 samples for a single MiSeq run.

Table S3: Distribution HIV Subtypes Successfully Sequenced

Subtype	Both Methods	Sanger Only	MiSeq Only
A1	164 (19.7%)	10 (20.4%)	10 (16.7%)
B	458 (55%)	27 (55.1%)	28 (46.7%)
C	62 (7.5%)	1 (2%)	7 (11.7%)
D	100 (12%)	7 (14.3%)	10 (16.7%)
G	9 (1.1%)	0 (0%)	1 (1.7%)
H	4 (0.5%)	0 (0%)	0 (0%)
AE	12 (1.4%)	1 (2%)	0 (0%)
Recombinant	6 (0.7%)	1 (2%)	0 (0%)
Undetermined	17 (2%)	2 (4.1%)	4 (6.7%)

HIV subtyping was performed using RIP using a 90% confidence threshold and a 200-bp window size (http://www.hiv.lanl.gov/content/sequence/RIP/RIP.html)

Table S4: Prevalence of NNRTI and NRTI Drug Resistance Mutations Detected by MiSeq and Sanger Sequencing

Drug Class	Codon	Amino Acid	Canada (n=546)			UARTO (n=286)		
			MiSeq (5% Mixture)	MiSeq (20% Mixture)	Sanger	MiSeq (5% Mixture)	MiSeq (20% Mixture)	Sanger
NNRTI	100	I	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	101	E	9 (1.6%)	6 (1.1%)	7 (1.3%)	3 (0.5%)	3 (0.5%)	3 (0.5%)
NNRTI	101	H	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	101	P	1 (0.2%)	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	103	N	48 (8.8%)	39 (7.1%)	39 (7.1%)	1 (0.2%)	1 (0.2%)	1 (0.2%)
NNRTI	106	A	1 (0.2%)	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	106	M	3 (0.5%)	2 (0.4%)	2 (0.4%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	108	I	15 (2.7%)	10 (1.8%)	12 (2.2%)	3 (0.5%)	0 (0%)	2 (0.4%)
NNRTI	138	A	25 (4.6%)	19 (3.5%)	18 (3.3%)	25 (4.6%)	20 (3.7%)	18 (3.3%)
NNRTI	138	G	4 (0.7%)	3 (0.5%)	3 (0.5%)	1 (0.2%)	0 (0%)	0 (0%)
NNRTI	138	K	8 (1.5%)	4 (0.7%)	3 (0.5%)	2 (0.4%)	1 (0.2%)	1 (0.2%)
NNRTI	138	Q	1 (0.2%)	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	138	R	1 (0.2%)	1 (0.2%)	0 (0%)	1 (0.2%)	1 (0.2%)	1 (0.2%)
NNRTI	181	C	14 (2.6%)	13 (2.4%)	13 (2.4%)	1 (0.2%)	1 (0.2%)	1 (0.2%)
NNRTI	181	I	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	181	V	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	188	C	0 (0%)	0 (0%)	0 (0%)	1 (0.2%)	1 (0.2%)	1 (0.2%)
NNRTI	188	H	2 (0.4%)	0 (0%)	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)
NNRTI	188	L	1 (0.2%)	1 (0.2%)	1 (0.2%)	1 (0.2%)	1 (0.2%)	1 (0.2%)
NNRTI	190	A	10 (1.8%)	10 (1.8%)	10 (1.8%)	1 (0.2%)	0 (0%)	0 (0%)
NNRTI	190	S	0 (0%)	0 (0%)	0 (0%)	2 (0.4%)	2 (0.4%)	2 (0.4%)
NNRTI	225	H	3 (0.5%)	2 (0.4%)	2 (0.4%)	0 (0%)	0 (0%)	0 (0%)
NNRTI	230	L	2 (0.4%)	2 (0.4%)	2 (0.4%)	0 (0%)	0 (0%)	0 (0%)
NRTI	115	F	1 (0.2%)	1 (0.2%)	1 (0.2%)	3 (0.5%)	2 (0.4%)	2 (0.4%)

Drug Class	Codon	Amino Acid	Canada (n=546) MiSeq (5% Mixture)	Canada (n=546) MiSeq (20% Mixture)	Canada (n=546) Sanger	UARTO (n=286) MiSeq (5% Mixture)	UARTO (n=286) MiSeq (20% Mixture)	UARTO (n=286) Sanger
NRTI	116	Y	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NRTI	151	M	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
NRTI	184	I	15 (2.7%)	11 (2%)	12 (2.2%)	3 (0.5%)	3 (0.5%)	3 (0.5%)
NRTI	184	V	37 (6.8%)	36 (6.6%)	38 (7%)	6 (1.1%)	5 (0.9%)	4 (0.7%)
NRTI	210	W	16 (2.9%)	11 (2%)	7 (1.3%)	1 (0.2%)	0 (0%)	0 (0%)
NRTI	215	F	1 (0.2%)	1 (0.2%)	2 (0.4%)	0 (0%)	0 (0%)	0 (0%)
NRTI	215	Y	9 (1.6%)	8 (1.5%)	7 (1.3%)	0 (0%)	0 (0%)	0 (0%)
NRTI	219	E	4 (0.7%)	4 (0.7%)	3 (0.5%)	2 (0.4%)	0 (0%)	1 (0.2%)
NRTI	219	Q	1 (0.2%)	1 (0.2%)	1 (0.2%)	0 (0%)	0 (0%)	0 (0%)

Number and frequency (%) of sequences containing HIV drug resistance mutations as detected by MiSeq and Sanger in n=832 samples successfully sequenced by both methods. Consensus sequences were generated from mapped MiSeq reads with mixed bases called when minority nucleotides were observed in at least 5% (5% Mixture) or 20% (20% Mixture) of sequence reads.

Table S5: Effect of Read Coverage Levels on Sensitivity and Specificity of MiSeq in Detecting Resistance Mutations Observed by Sanger Sequencing

Minimum Coverage (# Reads)	NNRTI		NRTI	
	Sensitivity	Specificity	Sensitivity	Specificity
0	96.75%	98.68%	90.91%	99.14%
5	96.75%	98.94%	90.91%	99.14%
10	96.72%	99.33%	93.75%	99.50%
15	98.33%	99.46%	93.75%	99.62%
20	98.33%	99.46%	93.75%	99.62%
25	98.33%	99.45%	96.77%	99.62%
30	98.33%	99.45%	96.77%	99.62%
35	98.33%	99.45%	96.77%	99.62%
40	98.33%	99.45%	96.77%	99.62%
45	98.33%	99.45%	96.77%	99.62%
50	98.32%	99.58%	96.77%	99.61%
55	98.32%	99.58%	96.77%	99.61%
60	98.32%	99.58%	96.77%	99.61%
65	98.32%	99.58%	96.77%	99.61%
70	98.32%	99.58%	96.77%	99.61%
75	98.32%	99.58%	96.77%	99.61%
80	98.32%	99.58%	96.77%	99.61%
85	98.32%	99.58%	96.77%	99.61%
90	98.32%	99.58%	96.77%	99.61%
95	98.32%	99.58%	96.77%	99.61%
100	98.32%	99.58%	96.77%	99.61%
105	98.31%	99.58%	96.72%	99.61%
110	98.31%	99.58%	96.72%	99.61%
115	98.31%	99.58%	96.72%	99.61%
120	98.31%	99.58%	96.72%	99.61%
125	98.31%	99.58%	96.72%	99.61%
130	98.31%	99.58%	96.72%	99.61%
135	98.31%	99.58%	96.72%	99.61%
140	98.31%	99.58%	96.72%	99.61%
145	98.31%	99.58%	96.72%	99.61%

Minimum Coverage (# Reads)	NNRTI		NRTI	
	Sensitivity	Specificity	Sensitivity	Specificity
150	98.31%	99.58%	96.72%	99.61%
160	98.31%	99.58%	96.72%	99.61%
170	98.31%	99.58%	96.72%	99.61%
180	98.29%	99.58%	96.67%	99.61%
190	98.29%	99.58%	96.67%	99.61%
200	98.29%	99.57%	96.67%	99.61%
250	98.28%	99.57%	96.67%	99.60%
300	98.28%	99.57%	96.67%	99.73%
350	98.26%	99.57%	96.67%	99.87%
400	98.26%	99.57%	96.67%	99.87%
450	98.26%	99.57%	96.67%	99.87%
500	98.25%	99.57%	96.61%	99.87%
550	98.23%	99.56%	96.61%	99.87%
600	98.21%	99.56%	96.61%	99.87%
650	98.21%	99.56%	96.61%	99.86%
700	98.20%	99.56%	96.55%	99.86%
750	98.20%	99.56%	96.55%	99.86%
800	98.20%	99.55%	96.55%	99.86%
850	98.20%	99.55%	96.55%	99.86%
900	98.20%	99.55%	96.55%	99.86%
950	98.18%	99.55%	96.55%	99.86%
1000	98.18%	99.54%	96.55%	99.86%

Nucleotide mixtures were called when minority bases were observed in at least 20% of MiSeq sequence reads.

Figure S1: Sanger and MiSeq Sequencing Success Rate Stratified by Plasma Viral Load and Cohort

Overall, 881 (80%) and 892 (81%) clinical samples were successfully sequenced by the Sanger and MiSeq methods, respectively, with 832 (75%) having sequences from both methods. Sequencing failure rate was driven largely by sample pVL without any obvious amplification bias in either cohort. Numbers above bars represent the total number of samples tested in each pVL category for each cohort.

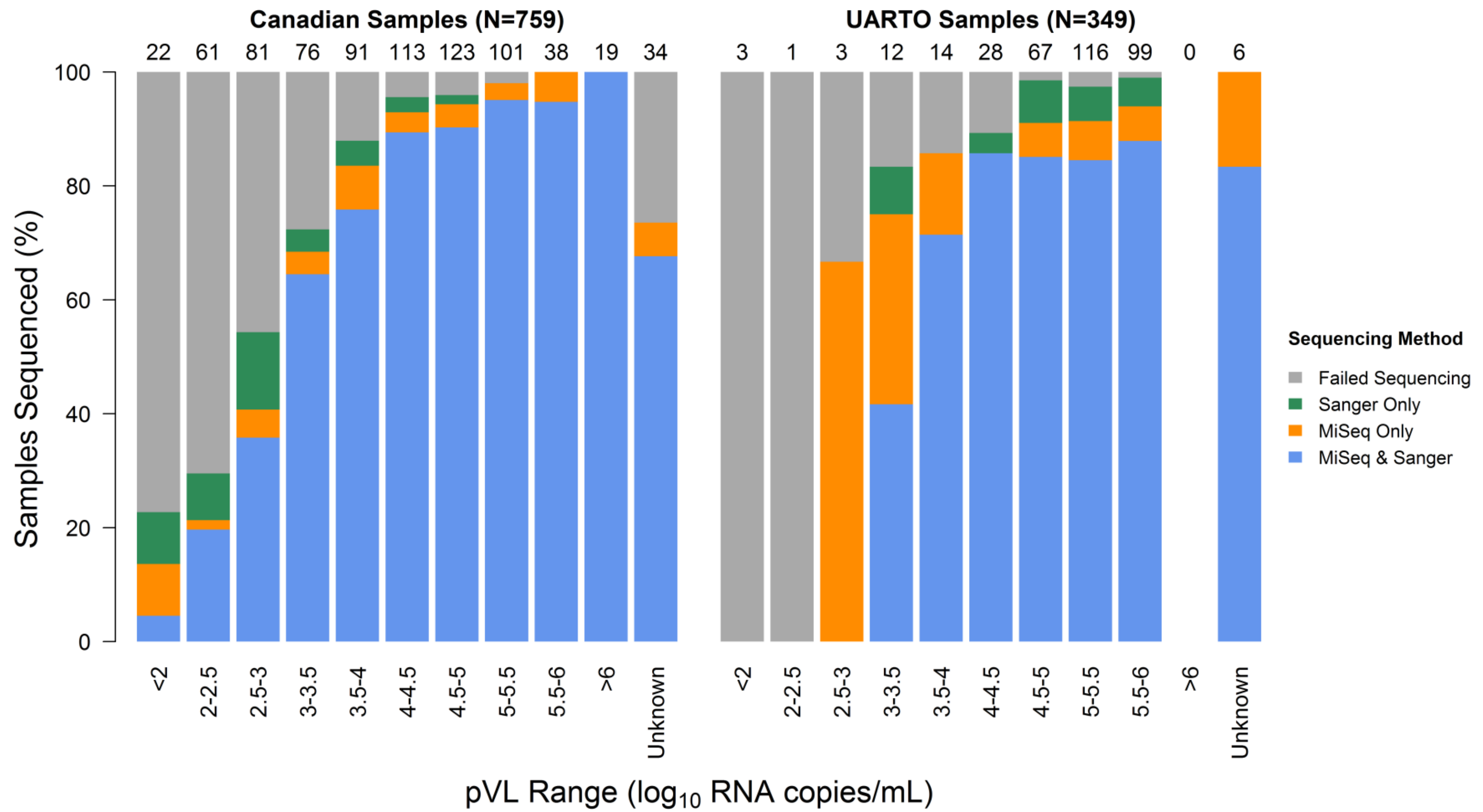

Figure S2: Proportion of Samples with Detectable Resistance Mutations with Varying Mixture Calling Thresholds

Drug resistance interpretations of clinical samples (N=832) successfully sequenced by Sanger and MiSeq methods. Orange and blue bars represent the proportion of samples with observed NNRTI and NRTI resistance (≥ 1 mutation) from MiSeq analysis. The dashed and dotted lines represent the results from Sanger analysis for NNRTI and NRTI resistance, respectively. The effect of varying the MiSeq nucleotide mixture calling threshold suggests that the sequence coverage obtained may be sufficient to identify lower-frequency resistance mutations.

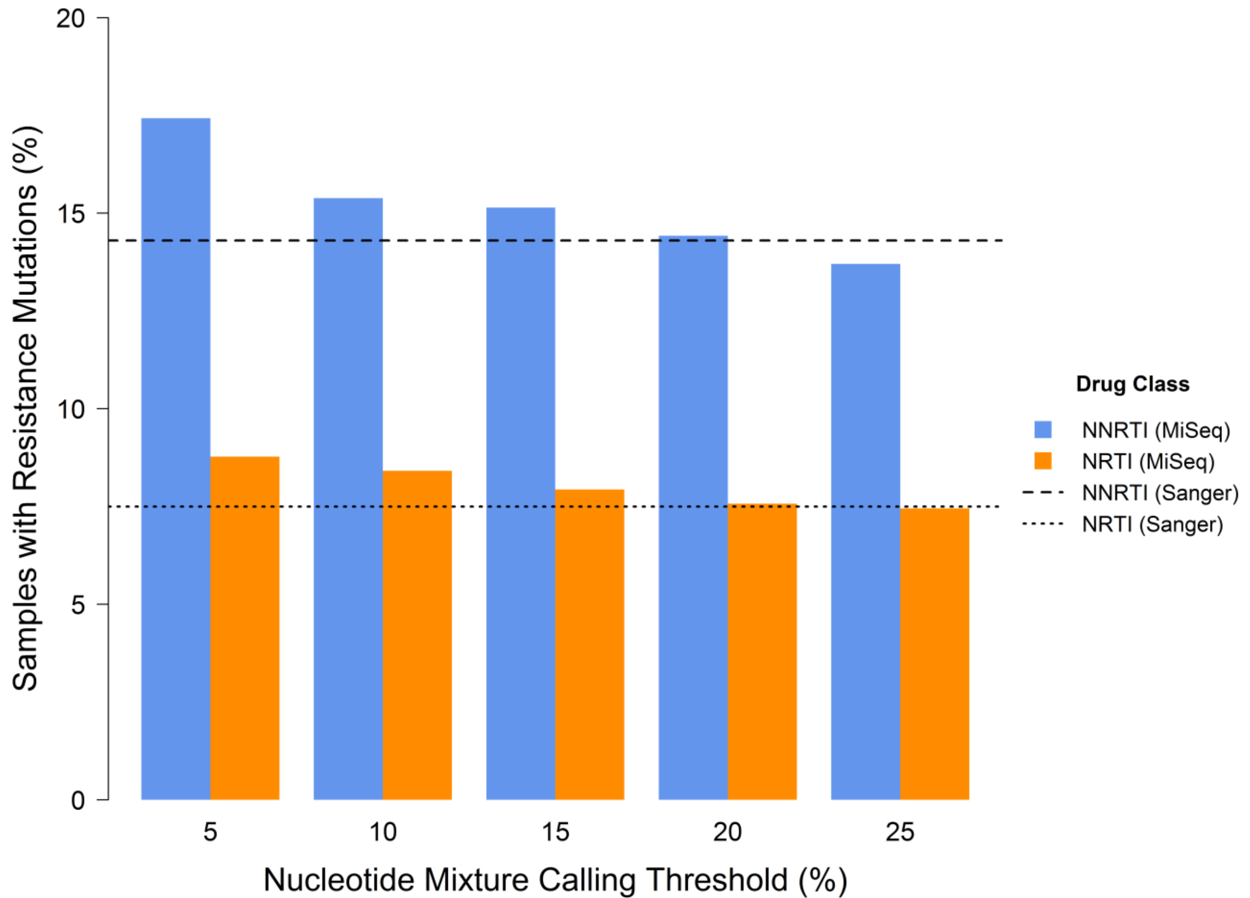

Figure S3: Sanger and MiSeq Nucleotide Sequence Concordance Stratified by Plasma Viral Load and Cohort

Sequence concordance between Sanger and MiSeq sequencing was high across all viral load strata regardless of the sample country of origin. Outliers beyond 1.5 IQR of the box hinge, represented by dots, are due largely to high numbers of mixed base calls in selected MiSeq sequences. Clinical samples without viral load data ("Unknown") were also successfully sequenced by both methods and yielded generally concordant results. Numbers above boxes represent the total number of successfully sequenced by both methods in each pVL category for each cohort.