Sensitive HIV Drug Resistance NGS Testing with and without Unique Molecular Identifiers

Kevin McCormick[1], Kerri Penrose[1], Rahil Sethi[2], Jacob Waldman[2], Uma Chandran[2], John Mellors[1] and Urvi Parikh[1]

[1]University of Pittsburgh - School of Medicine - Department of Infectious Disease

[2]University of Pittsburgh - School of Medicine - Department of Biomedical Bioinformatics

Background

- Sensitive next-generation sequencing (NGS) HIV drug resistance (HIV DR) surveillance is needed for optimal future HIV treatment regiments (Fig 1).

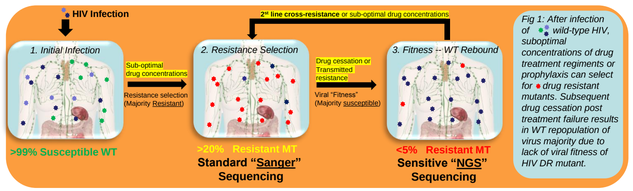


Fig 1: After infection of ● wild-type HIV, suboptimal concentrations of drug treatment regiments or prophylaxis can select for ● drug resistant mutants. Subsequent drug cessation post treatment failure results in WT repopulation of virus majority due to lack of viral fitness of HIV DR mutant.

- Adding a 10bp Unique Molecular Identifier (UMI) sequence to each HIV template in HIV DR NGS reduces polymerase fidelity error, PCR bias and sequence recombination.[1,2]
- However, the UMI-based NGS library preparation and pipeline processing is complex and inefficient (~3% UMI sequence yield to starting templates), limiting its use for HIV DR monitoring in low-middle income countries.
- A comprehensive study is urgently needed to demonstrate the limit sampling depth of a non-UMI assay to show that the added confidence of a UMI-assays outweighs the cost of preforming a more complex test.
- Preliminary data suggested that there is no difference in the sensitivities of the two assay (Table 1).

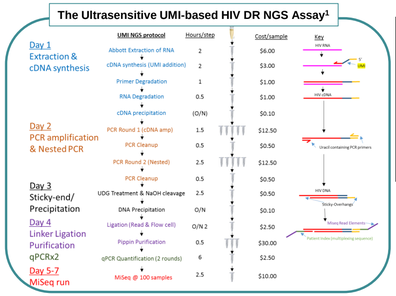
Objective

- To perform an assessment of the 3 major steps likely to influence the sensitivity of a UMI-based verses a non UMI-based assay: 1. Inefficient RT using primers with long UMI overhangs (Panels 2 & 3), 2. Skewed allelic amplification from PCR bias (Panel 4), 3. PCR recombination from MiSeq adapter addition (Panel 5) to determine if more affordable approaches could be employed to overcome limitations from non-UMI NGS artifacts.

Materials & Methods

- We used a previously characterized wild-type:drug-resistant HIV mixture panel (Panel 1-5) and a dataset of 33 HIV positive clinical samples (Panel 6) to determine the sensitivity of detecting minority HIV drug resistance templates with and without UMI-based consensus building.

The Ultrasensitive UMI-based HIV DR NGS Assay[1]



UMI NGS Protocol

High Sensitivity
- <1% low copy

High Cost
- 83.00/sample

Low Throughput
- 2 hours/sample
- 50 samples/week

High Complexity
- 15 steps
- Abbott, Real-time, Miseq, Pippin
- Complex sequence analysis

Non-UMI Protocol

Low Sensitivity
- 5% low frequency

Low Cost
- 19.24/sample

High Throughput
- >500 samples/week
- 6 min/sample (@ 100 samples/run)

Low Complexity
- Paseq.org software
- 8 steps

Results

1. Sensitive detection of low frequency HIV DR mutations using a WT:MT virus mixture panel with/without UMI consensus building.

Expected frequency of MT virus	100.0%	20.0%	10.0%	5.0%	2.5%	1.0%	0.5%	0.0%
UMI NGS assay	100.0%	20.3%	8.3%	3.7%	1.6%	0.5%	0.3%	0.0%
non-UMI NGS assay	100.0%	26.0%	14.0%	7.3%	3.3%	1.2%	0.5%	0.0%

2: Increased concentrations of RT primer improves limited RT efficiency in samples with polymorphic primer binding sites.

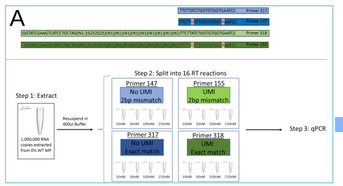
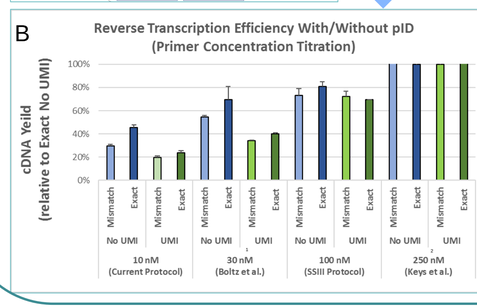


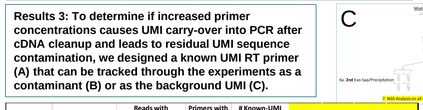
Fig 1: A. Experimental design showing the primers used (primer polymorphisms in RED) to reverse transcribe HIV RNA B. cDNA yield determined by quantitative PCR.



3: Increased concentrations of RT UMI primer (above) are not represented at levels to cause skewing of NGS data analysis.

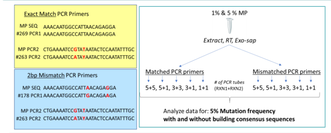
[Primer] nM	Cleanup Method	Reads with consensus	Primers with consensus	# Known-UMI Reads	% Known-UMI Reads/Total Reads
250	Exo-Sap (NCI)	100356	1397	96384	0.05%
100	Exo-Sap (NCI)	107920	1497	173426	0.03%
30	Exo-Sap (NCI)	103626	1258	260258	0.01%
10	Exo-Sap (NCI)	103988	1076	329151	0.00%
250	Beads (Swanstrom)	32198	3359	84918	0.28%
100	Beads (Swanstrom)	54181	7801	88189	0.05%
30	Beads (Swanstrom)	63008	8818	120052	0.01%
10	Beads (Swanstrom)	72931	9137	65321	0.01%

Results 3: To determine if increased primer concentrations causes UMI carry-over into PCR after cDNA cleanup and leads to residual UMI sequence contamination, we designed a known UMI RT primer (A) that can be tracked through the experiments as a contaminant (B) or as the background UMI (C).



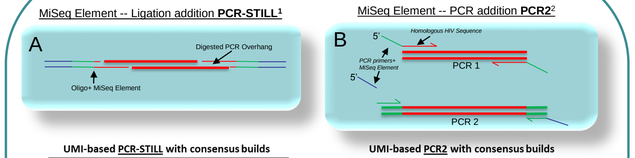
[Primer] nM	Cleanup Method	Reads with consensus	Primers with consensus	# Known-UMI Reads	# Unknown UMI with ≥ 3 reads	
250	Exo-Sap (NCI)	96384	1	96384	2	AAAATCATTC, TTATGCGTTC
100	Exo-Sap (NCI)	173426	1	173426	1	GGTGTGCAAG
30	Exo-Sap (NCI)	260258	1	260258	2	ATCGCTCTAT, GTATGCGCAC
10	Exo-Sap (NCI)	329151	1	329151	0	
250	Beads (Swanstrom)	84918	1	84918	2	ATCGCTCGAG, CATTAATACT
100	Beads (Swanstrom)	88189	1	88189	0	
30	Beads (Swanstrom)	120052	1	120052	1	CGAGCAATAC
10	Beads (Swanstrom)	65321	1	65321	1	GCTCGATCTC

4: Splitting samples into multiple PCR reactions does not improve the PCR bias in reactions where primers contain polymorphic bases (red circle).



Sample	# of PCR1 primer binding site mismatches	# of Reactions PCR 1	# of Reactions PCR 2	Reads with consensus	Primers with consensus	Expected%	UMI Actual%	Non-UMI Actual%
1	0	5	5	18142	1447	1%	0.9%	1.5%
2	0	5	5	19786	1213	5%	4.0%	7.0%
3	2	5	5	18057	248	1%	0.8%	0.0%
4	2	5	5	16919	215	5%	10.7%	7.7%
5	0	5	1	17954	1244	1%	1.5%	1.7%
6	0	5	1	19090	1170	5%	6.5%	7.6%
7	2	5	1	12330	120	1%	3.3%	1.9%
8	2	5	1	13703	178	5%	11.2%	8.3%
9	0	3	3	16816	1168	1%	0.9%	1.5%
10	0	3	3	16309	1025	5%	4.6%	7.5%
11	2	3	3	17346	246	1%	1.2%	1.2%
12	2	3	3	14571	183	5%	8.2%	11.4%
13	0	3	1	15347	1092	1%	1.5%	1.7%
14	0	3	1	15832	989	5%	7.3%	8.0%
15	2	3	1	12873	200	1%	1.0%	0.0%
16	2	3	1	15881	184	5%	6.0%	8.0%
17	0	1	1	23977	1036	1%	1.4%	1.6%
18	0	1	1	25168	911	5%	8.5%	9.5%
19	2	1	1	19290	253	1%	0.0%	0.0%
20	2	1	1	18891	211	5%	8.5%	10.3%

5: PCR Recombination from the PCR addition of MiSeq elements does not cause recombination or skew NGS data.



UMI-based PCR-STILL with consensus builds

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
20.0%	20.3%	20.3%	20.3%	20.3%	20.3%
10.0%	8.3%	8.3%	8.3%	8.3%	8.3%
5.0%	3.7%	3.7%	3.7%	3.7%	3.7%
2.5%	1.6%	1.6%	1.6%	1.6%	1.6%
1.0%	0.5%	0.5%	0.5%	0.5%	0.5%
0.5%	0.3%	0.3%	0.3%	0.3%	0.3%
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

UMI-based PCR2 with consensus builds

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	100.0%	*	100.0%	100.0%	100.0%
20.0%	20.2%	*	20.4%	29.0%	23.2%
10.0%	5.2%	*	6.0%	15.8%	9.0%
5.0%	2.2%	*	2.9%	7.9%	4.3%
2.5%	0.9%	*	1.4%	4.1%	2.1%
1.0%	0.2%	*	0.2%	0.2%	0.2%
0.5%	0.2%	*	0.2%	0.2%	0.2%
0.0%	0.0%	*	0.0%	0.0%	0.0%

UMI-based PCR-STILL without consensus builds

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
20.0%	26.0%	26.0%	26.0%	26.0%	26.0%
10.0%	14.0%	14.0%	14.0%	14.0%	14.0%
5.0%	7.3%	7.3%	7.3%	7.3%	7.3%
2.5%	3.3%	3.3%	3.3%	3.3%	3.3%
1.0%	1.1%	1.2%	1.2%	1.3%	1.2%
0.5%	0.5%	0.5%	0.5%	0.6%	0.5%
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

UMI-based PCR2 without consensus builds

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	99.7%	*	99.7%	99.8%	99.7%
20.0%	28.5%	*	29.6%	29.5%	29.2%
10.0%	13.9%	*	14.2%	14.1%	14.1%
5.0%	6.9%	*	7.4%	7.3%	7.2%
2.5%	3.7%	*	4.3%	4.2%	4.1%
1.0%	1.1%	*	1.2%	1.2%	1.2%
0.5%	1.1%	*	1.2%	1.2%	1.2%
0.0%	0.0%	*	0.0%	0.0%	0.0%

Non UMI-based PCR-STILL

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	100.0%	100.0%	100.0%	100.0%	100.0%
20.0%	28.0%	28.0%	27.0%	27.0%	27.5%
10.0%	14.6%	14.6%	14.6%	14.6%	14.6%
5.0%	7.5%	7.6%	7.8%	7.8%	7.7%
2.5%	3.4%	3.4%	3.5%	3.5%	3.5%
1.0%	1.3%	1.3%	1.5%	1.5%	1.4%
0.5%	0.7%	0.6%	0.6%	0.6%	0.6%
0.0%	0.0%	0.0%	0.0%	0.0%	0.0%

Non UMI-based PCR2

Expecte	K65R	K103N	Y181C	M184V	Average
100.0%	100.0%	*	100.0%	100.0%	100.0%
20.0%	27.9%	*	29.0%	29.0%	28.6%
10.0%	14.9%	*	15.8%	15.8%	15.5%
5.0%	7.8%	*	7.9%	7.9%	7.9%
2.5%	3.9%	*	4.1%	4.1%	4.0%
1.0%	1.3%	*	1.5%	1.5%	1.4%
0.5%	0.6%	*	0.8%	0.8%	0.7%
0.0%	0.0%	*	0.0%	0.0%	0.0%

Clinical Results

6: The use of UMI in clinical samples improves the assay specificity but does not increase the overall sensitivity of detection.

Patient Sample	Reads with consensus	Primers with consensus	UMI-Based HIV DR NGS	non UMI-Based HIV DR NGS
1	138008	18293	0.6% D192N, 0.5% E204K	**10.25 R83S**, 0.5% E122K, 0.51% E194K, 0.7% D192N, 0.73% D204K, **2.7% H19Q**
2	189026	2723	98.5 Q174S/1.5% Q174N	96.3 Q174S/2.4% Q174N/**1.4% Q174R**
3	123396	17480	65.7% W88C, 12.6% K102R, 1.2% K104R, 75.4% V118M, 65.6% S122N, 32% S123D, 10.1% S162Q/23.6% S162F/2.7% S162A, 65.8% T165I, 10.2% A173T, 23.8% E177D, 76.0% I202V, 0.8% R206K	65.8% W88C, 12.7% 102R, 1.22% K104R, 1.1% V118M, 75.5% V118M, 66.8% G123N, 31.4% G123D 11.2% S123G, 10.0% S162C, 11.0% S162A, 22.4% S162F, 66.8% T165I 10% A173T, 22.0% E177D, 0.6% G196E, 76.0% I202V, 0.6% E203K, 1.0% R206K,
4	184327	968	0.5% R143K, 0.5% V179I, 1.9% V189I, 0.6% E194K	**1.5% K101N**, 0.58% V118I, 0.8% R143K, 0.5% A173T, 0.53 I178V, 1.0% V179I, 1.9%V189I, 0.6% E194K, 0.58 D204K,
5	146041	1738	0.5% E194K	0.58% E194K, 0.52% T165I, 0.5% Q207A, **1.34% K101N, 4.1% Q91H**
6	157113	2284	0.9% E169K	0.6% A173T, 0.6% E203, 0.6% E122K, 1.27% E169K
7	185773	7065	0.8% Q151P, 20% A173T, 2.5% E177D, 2.5% I178L, 0.8% E194K	0.9% Q151K, 31.8% A173T, 6.09% E177D, 6.15% I178L, 0.6% G190E, 0.9% D192N, 1.55% E194K, 1.05% E203K, 3.4% D204K, 0.6% Q207K
8	147520	8091	1.5% K211R	0.7% Q207E, 0.8% D204K, 1.6% K211R
9	161785	8126	-	-
10	58198	9791	52.7% K102R, 55.4% K211Q	47.5 K102R, 0.5% Q207T, 54.9% K211Q
11	107074	11909	8.7% F87L, 0.6% E122A, Its odd that 204 is only 98%	8.7% F87L, 3.1% D204E
12	184976	3504	-	**3.4% K103I**
13	202544	7478	-	0.5% F214L
14	248138	12600	-	-
15	264238	16143	-	-
16	132300	16725	43.4% I178L, 43.6% V179I/55.7% V179T, 54.4% Q191P, 51.5% E207D/16.3% E207G, 31.92 K211R	47.8% I178L, 47.9% V179I/50.4% V179T, 49.7% Q197P, 51.5% E207D/16.3% E207G, 31.92 K211R
17	115182	17057	4.0% S134N	6.1% S134N, 1.6% A173K, 1.7% K174Q 0.8% E194K
18	195257	25189	0.6% Q151K	0.6% V118I, 1.0% S123D, 0.5% S162N, 0.9% D204K
19	265361	25311	0.8% Q151P, 3.3% A200T,	0.5% K173D, 0.5% A173T, 4.0% A200T
20	276259	423	0.9% Q151P, 0.7% E194K, 0.9% A200K,	K101 0.9% G112E, 0.6% D113N, 0.9% V189I, 0.7% E203K, 2.83% K104I, 1.0% E194K, 1.1% A200K, 0.6% R206K
21	313290	525	0.6% E194K	0.8% A114T, 0.7% E138K, 0.5% E194K,
22	284272	945	-	0.6% S162N, 0.6% A200V
23	230637	2288	2.1% K102R	6.9% K102R
24	173888	2782	0.6% E122K, 33.8% I178V, 0.6% D192N, 0.7% E194K,	1.3% E122K, 0.5 Q151P, 33.9% I178V, 0.9% D192N, 0.8% E194K, 0.7% R206K
25	242014	9303	2.2% P176L	2.2% P176L, 3.9% Q207K, 0.7% Q207E <--Another one that's odd. How is it only 98%?
26	240955	14581	-	-
27	281884	20493	-	0.5% A173T, 1.0% Q207E
28	282593	21716	-	-
29	209454	21897	0.5% D194N, 0.7% E196K, 0.8% E206K, 1.3% R208K	0.5 R143K, 0.5% E169K, 0.7% D192N, 0.9% E194K, 0.6% E203K, 1.5% E206K
30	138305	3585	1.9% Q151K	-
31	32965	5743	-	0.57% Q207E, **1.51% K211R, 2.16% L193I <--double check**
32	251700	8919	5.2% E207D	5.7% E207D
33	201222	13727	0.6% E122K, 0.7% E204K	0.8% R199K, 0.9% E204K, 1.3% A200R

Key: Concordant Mutations -- Discordant Mutations

Conclusions

- We observed limited RT efficiency with our current library prep protocol--particularly in sequences with primer-site polymorphisms—which can be improved 5 fold by increasing primer concentrations without significant risk of residual UMI contamination (Panels 2 & 3).
- Biased PCR amplification limited the total number of sequenced HIV templates in samples with PCR primer site polymorphisms; however, UMI consensus building was able to overcome data skewing from PCR amplicon preference (Panel 4).
- The PCR addition of MiSeq primers reduces assay complexity and cost, but does not cause recombination (Panel 5).
- In clinical specimens, the non-UMI based had similar level of sensitivity (low false negative mutations), but has a significant amount of false positives (low specificity) (Panel 6).

References

1) Boltz, V. F., Rausch, J., Shao, W., Hattori, J., Luke, B., Maldarelli, F., ... Coffin, J. M. (2016). Ultrasensitive single-genome sequencing: Accurate, targeted, next generation sequencing of hiv-1 rna. Retrovirology, 13(1), 1–17. https://doi.org/10.1186/s12977-016-0321-6

2) Keys, J. R., Zhou, S., Anderson, J. A., Eron, J. J., Rackoff, L. A., Jabara, C., & Swanstrom, R. (2015). Primer ID Informs Next-Generation Sequencing Platforms and Reveals Preexisting Drug Resistance Mutations in the HIV-1 Reverse Transcriptase Coding Domain. AIDS Research and Human Retroviruses, 31(6), 658–668. https://doi.org/10.1089/aid.2014.0031