LTR Parser Results

Kevin McCormick

22 June, 2020

Overview

This report contains the results of attempted primer & LTR identification in the following FASTQ file:

/home/kevin/dev/ltrparser/testdata/demo/fastq.gz/hivTest5_u3.R2.fastq.gz

Blankson project, sample: GTSP3503-2 (U3)

Analysis of common primer-LTR pairs

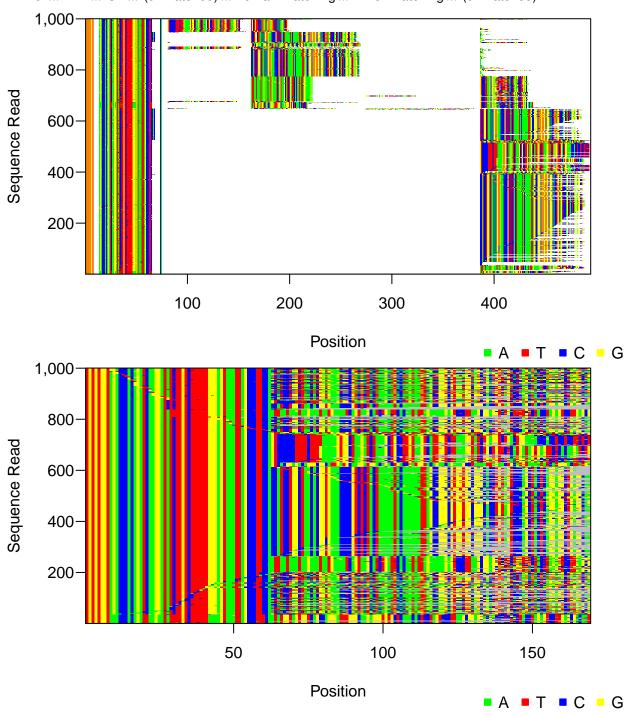
A total of 6 potential primer-LTR pairs were identified. This table shows how the reads captured by each primer-LTR pair compare to each other and to the raw FASTQ file.

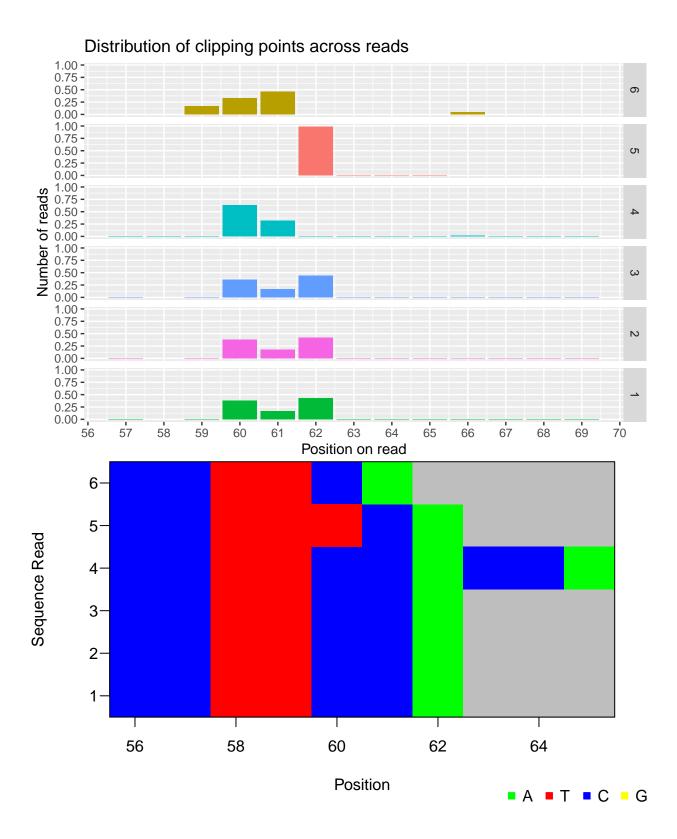
id	primer	LTR	expectedLTR	LANL
1	TGTGTGGT	AGACCCACAGATCAAGGATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA	FALSE	TRUE
2	TGTGTGGT	AGACCCACAGATCAAGGATCTCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA	FALSE	TRUE
3	TGTGTGGT	AGACCCACAGATCAAGGATATTTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA	FALSE	TRUE
4	TGTGTGGT	AGACCCACAGATCAAGGATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCACCA	FALSE	TRUE
5	TGTGTGGT	AAACCCACAGATTAAGAATATTTTGTCTTTTTTAAAAGTAAATTAGCCCTTTCA	FALSE	FALSE
6	TGTGTGGT	AGACCCACAGATCAAGGATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCA	FALSE	TRUE

Variable	Total	1	2	3	4	5	6
Distinct	8366	4320 (51.6%)	4143 (49.5%)	4119 (49.2%)	2572 (30.7%)	509 (6.1%)	96 (1.1%)
DistinctHuman	2589	1538 (59.4%)	1449 (56%)	1466 (56.6%)	607 (23.4%)	400 (15.4%)	34 (1.3%)
DistinctLinker	722	14 (1.9%)	14 (1.9%)	14 (1.9%)	0 (0%)	40 (5.5%)	0 (0%)
Max	1829	1829 (100%)	1829 (100%)	1829 (100%)	1829 (100%)	233 (12.7%)	14 (0.8%)
MaxHuman	1829	1829 (100%)	1829 (100%)	1829 (100%)	1829 (100%)	233 (12.7%)	14 (0.8%)
MaxLinker	12	2 (16.7%)	2 (16.7%)	2 (16.7%)	0 (0%)	12 (100%)	0 (0%)
Total	22628	14430 (63.8%)	14250 (63%)	14227 (62.9%)	9300 (41.1%)	1192 (5.3%)	132 (0.6%)
TotalHuman	7912	6088 (76.9%)	5996 (75.8%)	6014 (76%)	3132 (39.6%)	1036 (13.1%)	53 (0.7%)
TotalLinker	781	15 (1.9%)	15 (1.9%)	15 (1.9%)	0 (0%)	58 (7.4%)	0 (0%)

Visualization of reads

Reads are sectioned off into sections separated by white (nucleotide color legend is below the plot). Primer /// LTR /// CA /// (unmatched) /// human-matching /// linker-matching /// (unmatched)





logo	AGACCCACAGATCAAGGATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA	
6	AGACCCACAGATCAAGGATATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCA	53
5	AAACCCACAGATTAAGAATATTTTGTCTTTTTTAAAAGTAAATTAGCCCTTTCA	54
4	$\overline{A} \overline{G} A C C C A G \overline{G} A T \overline{G} \overline{C} T T T T T T T \overline{G} \overline{G} G G T A A A T T G G C C T \overline{C} C C C C$	57
3	AGACCCACAGATCAAGGATAT <mark>T</mark> TTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA	54
2	AGACCCACAGATCAAGGATCCTTTTTTTGGGAGTAAATTAGCCCTTCCA	54
1	$\mathtt{AGACCCACAGATCAAGGAT}$ $\overline{\mathtt{ATCTTGTCTTTTTTGGGAGTAAATTAGCCCTTCCA}$	54
HXB2	AGA <mark>T</mark> CCACAGATCAAGGATATCTTGTCTT <mark>C</mark> TTTGGGAGT <mark>G</mark> AATTAGCCCTTCCA	54
consensus	1*!*I!!!!!*!!*!!*!!*!!*!!*!!*!!***!!!***	

 \overline{X} non conserved \overline{X} $\geq 50\%$ conserved