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/hivTest7_u5.R2.fastq.gz

ODoherty project, sample: GTSP2314a (U5)
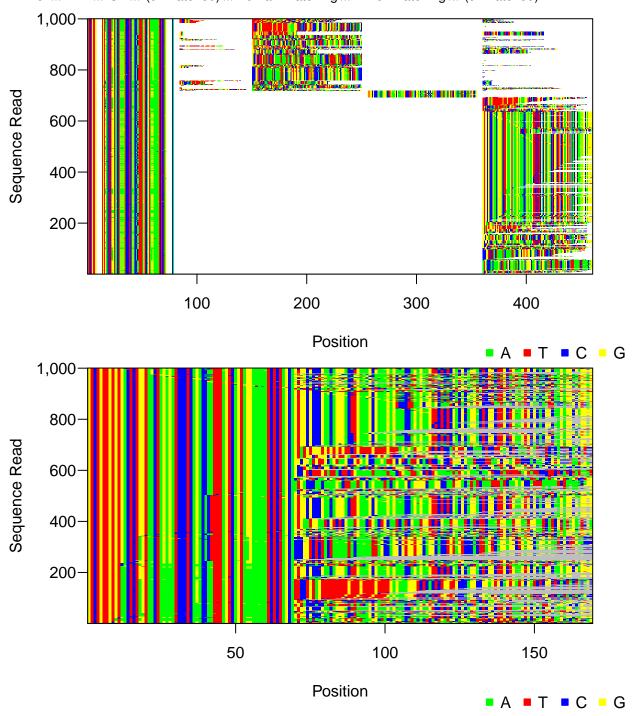
Analysis of common primer-LTR pairs

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.
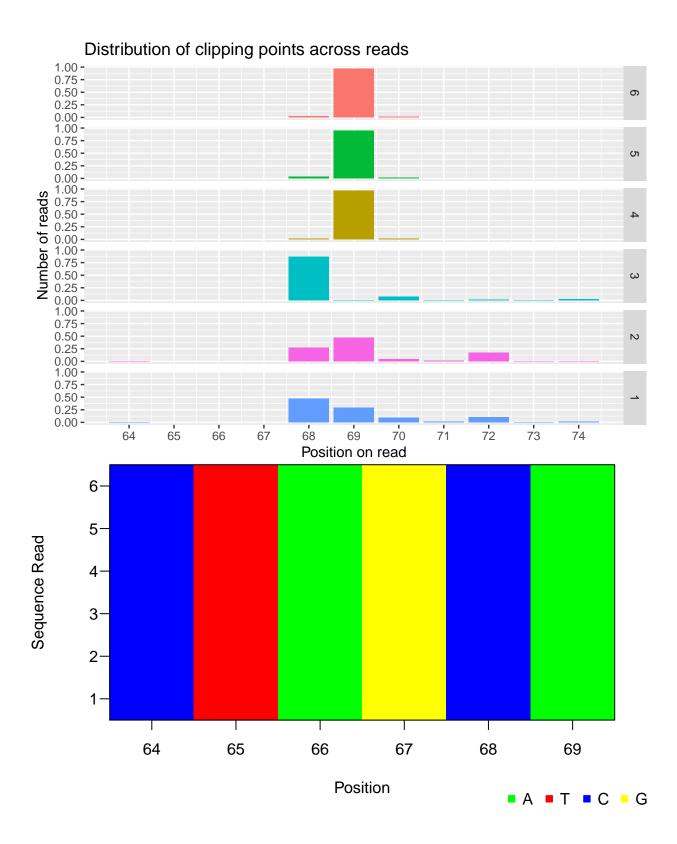
id	primer	LTR	expectedLTR	LANL
1	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
2	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
3	GTCTGTTG	TGTGACTCTAGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAAGAAAAATCTCTAGCA	FALSE	TRUE
4	GTCTGTTG	TGTAACTCTAGTAACTAAAAATCCCTCAGACCAATTTAGTCAGTAAAAAAAAATCTCTAGCA	FALSE	FALSE
5	GTCTGTTG	TGTAACTCTGGTAACTAAAAATCCCTCAAACCAATTTAGTCAGTAAAAAAAAATCTCTAGCA	FALSE	FALSE
6	GTCTGTTG	TATAACTCTAGTAACTAAAAATCCCTCAAACCAATTTAGTCAGTAAAAAAAAATCTCTAGCA	FALSE	FALSE

Variable	Total	1	2	3	4	5	6
Distinct	6443	2066 (32.1%)	2480 (38.5%)	652 (10.1%)	638 (9.9%)	563 (8.7%)	419 (6.5%)
DistinctHuman	1162	488 (42%)	486 (41.8%)	269 (23.1%)	245 (21.1%)	182 (15.7%)	190 (16.4%)
DistinctLinker	1180	73 (6.2%)	116 (9.8%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Max	1264	1264 (100%)	1264 (100%)	942 (74.5%)	405 (32%)	405 (32%)	405 (32%)
MaxHuman	942	942 (100%)	489 (51.9%)	942 (100%)	405 (43%)	405 (43%)	405 (43%)
MaxLinker	339	279 (82.3%)	279 (82.3%)	0 (0%)	0 (0%)	0 (0%)	0 (0%)
Total	23913	10130 (42.4%)	9958 (41.6%)	3324 (13.9%)	2127 (8.9%)	2028 (8.5%)	1462 (6.1%)
TotalHuman	4991	3172 (63.6%)	2184 (43.8%)	1740 (34.9%)	1106 (22.2%)	1019 (20.4%)	1028 (20.6%)
TotalLinker	4624	400 (8.7%)	460 (9.9%)	0 (0%)	0 (0%)	0 (0%)	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)

Distribution of clipping points across reads

```
logo       TGTGACTCTGGTAACTAGAGATCCCTCAGACCCAATTTAGTCAGTAAGAAAAATCTCTAGCA

6          TATAACTCTAGTAACTAAAAATCCCTCAAACCAATTTAGTCAGTAAAAAAAATCTCTAGCA    61
5          TGTAACTCTGGTAACTAAAAATCCCTCAAACCAATTTAGTCAGTAAAAAAAATCTCTAGCA    61
4          TGTAACTCTAGTAACTAAAAATCCCTCAGACCAATTTAGTCAGTAAAAAAAATCTCTAGCA    61
3          TGTGACTCTAGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAAGAAAAATCTCTAGCA    61
2          TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA    61
1          TGTGACTCTGGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAGGAAAAATCTCTAGCA    61
HXB2       TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA    61
consensus  !*!*!!!!!*!!!!!!!*!*!!!!!!!!!*!!***!!!!!!!!!!!!****!!!!!!!!!!!
```

⬚X non conserved

▣X ≥ 50% conserved