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/hivTest8_u5.R2.fastq.gz

ODoherty project, sample: GTSP2314b (U5)

Analysis of common primer-LTR pairs

A total of 9 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.
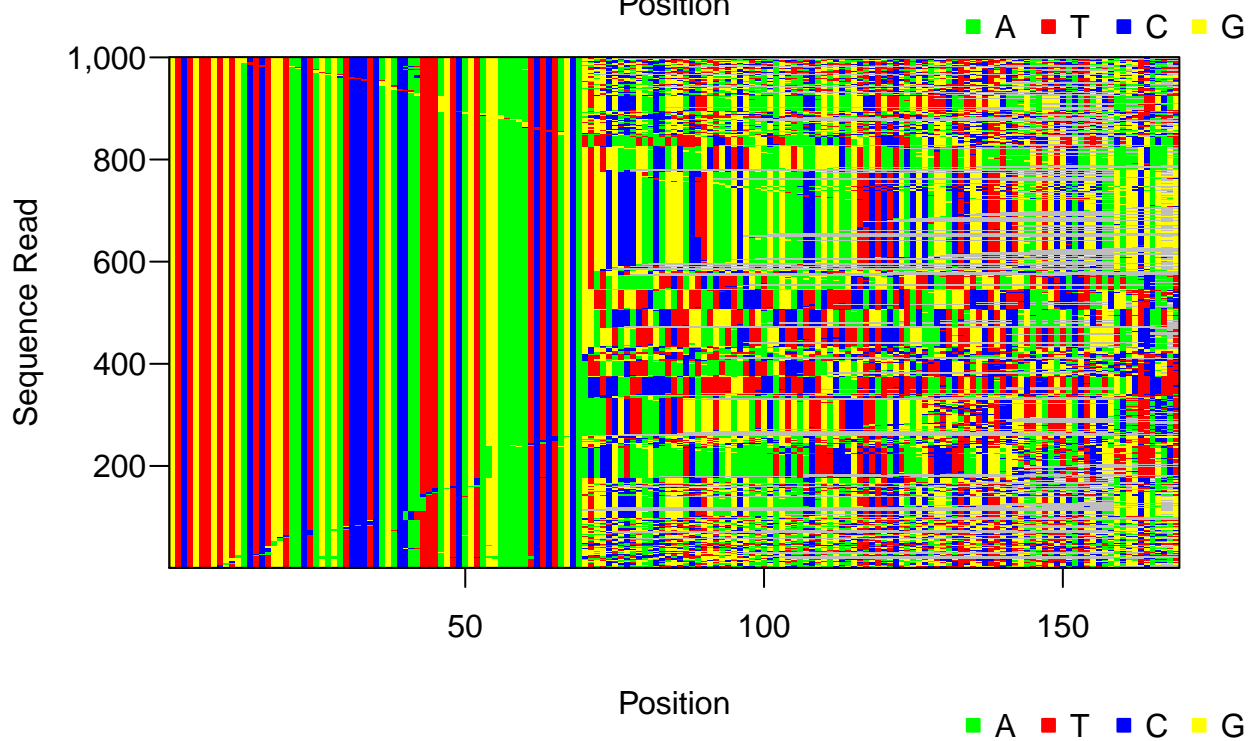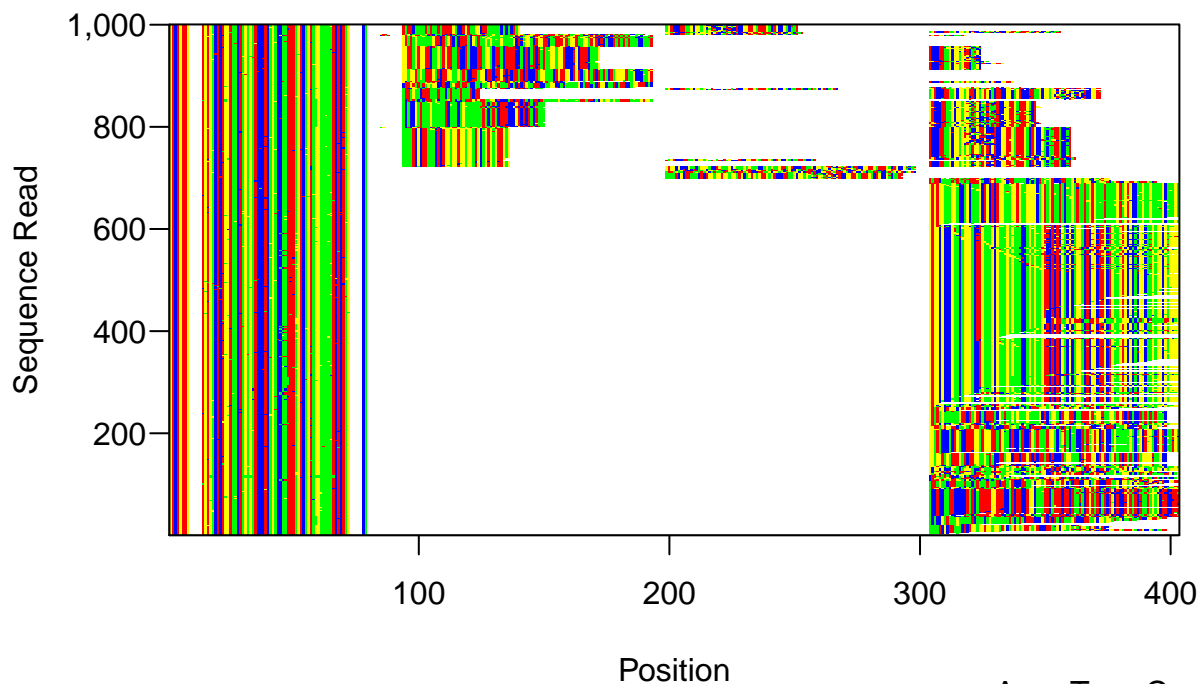
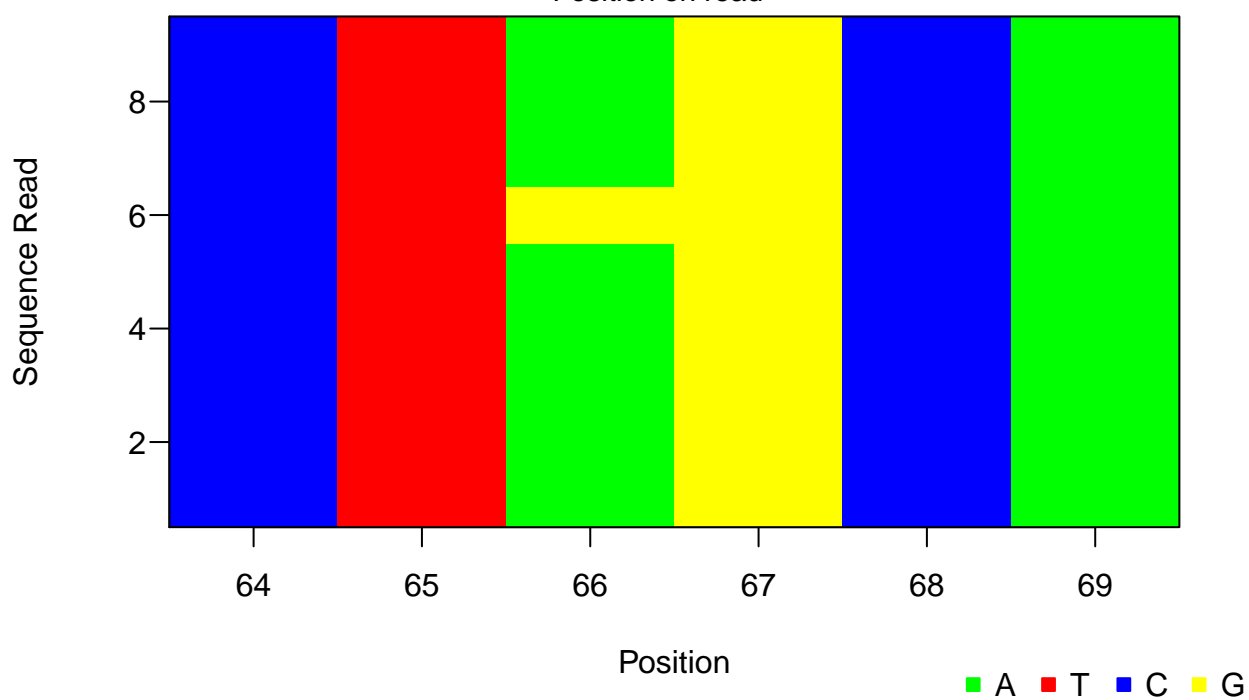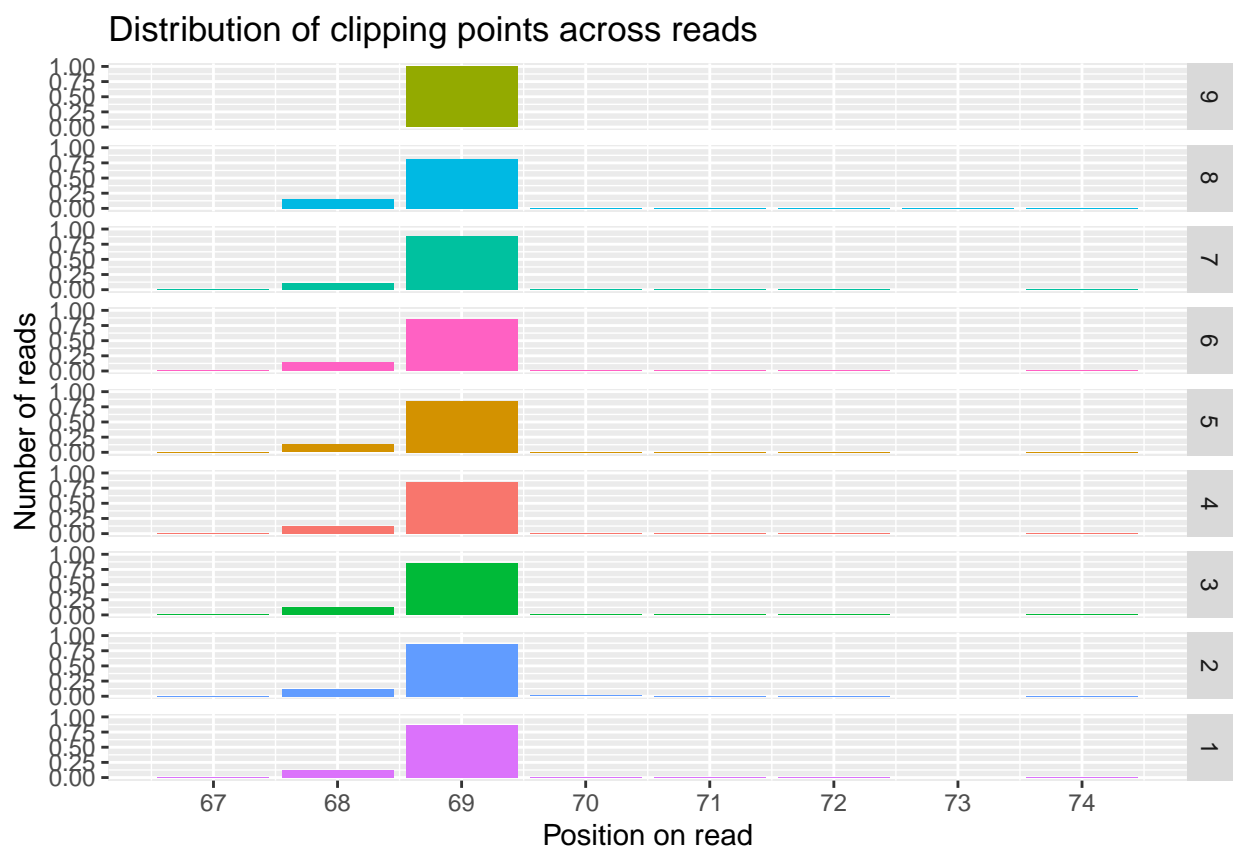| id	primer	LTR	expectedLTR	LANL
1	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
2	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAAGAAAAATCTCTAGCA	FALSE	TRUE
3	GTCTGTTG	TGTGACTCCGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
4	GTCTGTTG	TGCGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
5	GTCTGTTG	TGGGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
6	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTGGCA	FALSE	TRUE
7	GTCTGTTG	TGTGACTCTAGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAAGAAAAATCTCTAGCA	FALSE	TRUE
8	GTCTGTTG	TGTGACTCTGGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAGGAAAAATCTCTAGCA	FALSE	TRUE
9	GTCTGTTG	TGTAACTCTGGTAACTAAAAATCCCTCAGACCAATTTAGTCAGTAAAAAAAAACTCTAGCA	FALSE	FALSE

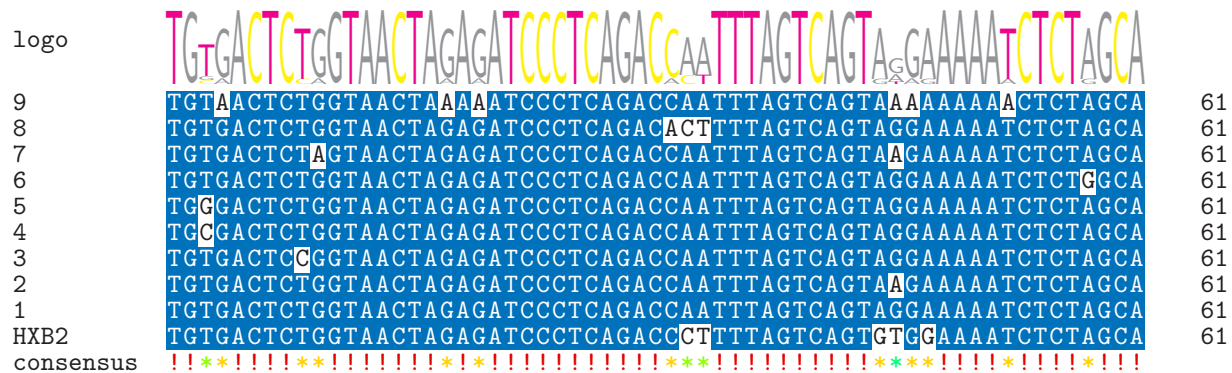| Variable	Total	1	2	3	4	5	6
Distinct	8947	4329 (48.4%)	3749 (41.9%)	3600 (40.2%)	3598 (40.2%)	3596 (40.2%)	3590 (40.1
DistinctHuman	1289	1149 (89.1%)	1139 (88.4%)	1017 (78.9%)	1018 (79%)	1017 (78.9%)	1012 (78.5
DistinctLinker	2419	183 (7.6%)	168 (6.9%)	168 (6.9%)	170 (7%)	170 (7%)	168 (6.9%)
Max	891	891 (100%)	891 (100%)	891 (100%)	891 (100%)	891 (100%)	891 (100%
MaxHuman	357	357 (100%)	357 (100%)	357 (100%)	357 (100%)	357 (100%)	357 (100%
MaxLinker	329	46 (14%)	46 (14%)	46 (14%)	46 (14%)	46 (14%)	46 (14%)
Total	23838	11482 (48.2%)	10195 (42.8%)	9959 (41.8%)	9958 (41.8%)	9954 (41.8%)	9948 (41.7
TotalHuman	2819	2672 (94.8%)	2660 (94.4%)	2470 (87.6%)	2472 (87.7%)	2471 (87.7%)	2466 (87.5
TotalLinker	6743	299 (4.4%)	284 (4.2%)	284 (4.2%)	286 (4.2%)	286 (4.2%)	284 (4.2%)

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     TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA

9        TGTAACTCTGGTAACTAAAAATCCCTCAGACCAATTTAGTCAGTAAAAAAAACTCTAGCA    61
8        TGTGACTCTGGTAACTAGAGATCCCTCAGACACTTTTAGTCAGTAGGAAAAATCTCTAGCA    61
7        TGTGACTCTAGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAAGAAAAATCTCTAGCA    61
6        TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTGGCA    61
5        TGGGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA    61
4        TGCGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA    61
3        TGTGACTCCGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA    61
2        TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAAGAAAATCTCTAGCA    61
1        TGTGACTCTGGTAACTAGAGATCCCTCAGACCAATTTAGTCAGTAGGAAAAATCTCTAGCA    61
HXB2     TGTGACTCTGGTAACTAGAGATCCCTCAGACCCTTTTAGTCAGTGTGGAAAATCTCTAGCA    61

consensus !!**!!!!**!!!!!!!*!*!!!!!!!!!!!!!***!!!!!!!!!!***!*!!!!*!!!!*!!!
```

X non conserved
X ≥ 50% conserved