

# LTR Parser Results

Kevin McCormick

01 May, 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/hivWater.fastq.gz

## Summary Statistics for raw FASTQ file

Total	411
Distinct	411
Max	1
TotalLinker	26
DistinctLinker	26
MaxLinker	1
TotalHuman	20
DistinctHuman	20
MaxHuman	1

## Explanation of terms

Total: Number of reads

Distinct: Number of distinct sequences

Max: Number of copies of most abundant sequence

TotalLinker: Number of reads containing R1 linker sequence

DisctinctLinker: Number of distinct sequences containing R1 linker sequence

MaxLinker: Number of copies of most abundant sequence containing R1 linker sequence

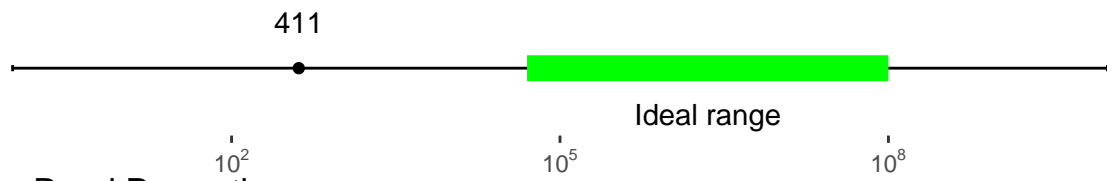
TotalHuman: Number of reads mapping to human genome

DisctinctHuman: Number of distinct sequences mapping to human genome

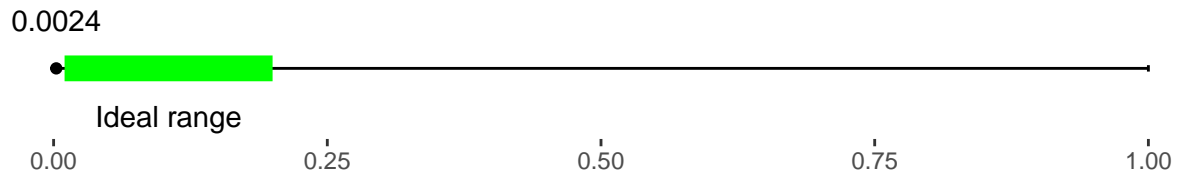
MaxHuman: Number of copies of most abundant sequence mapping to human genome

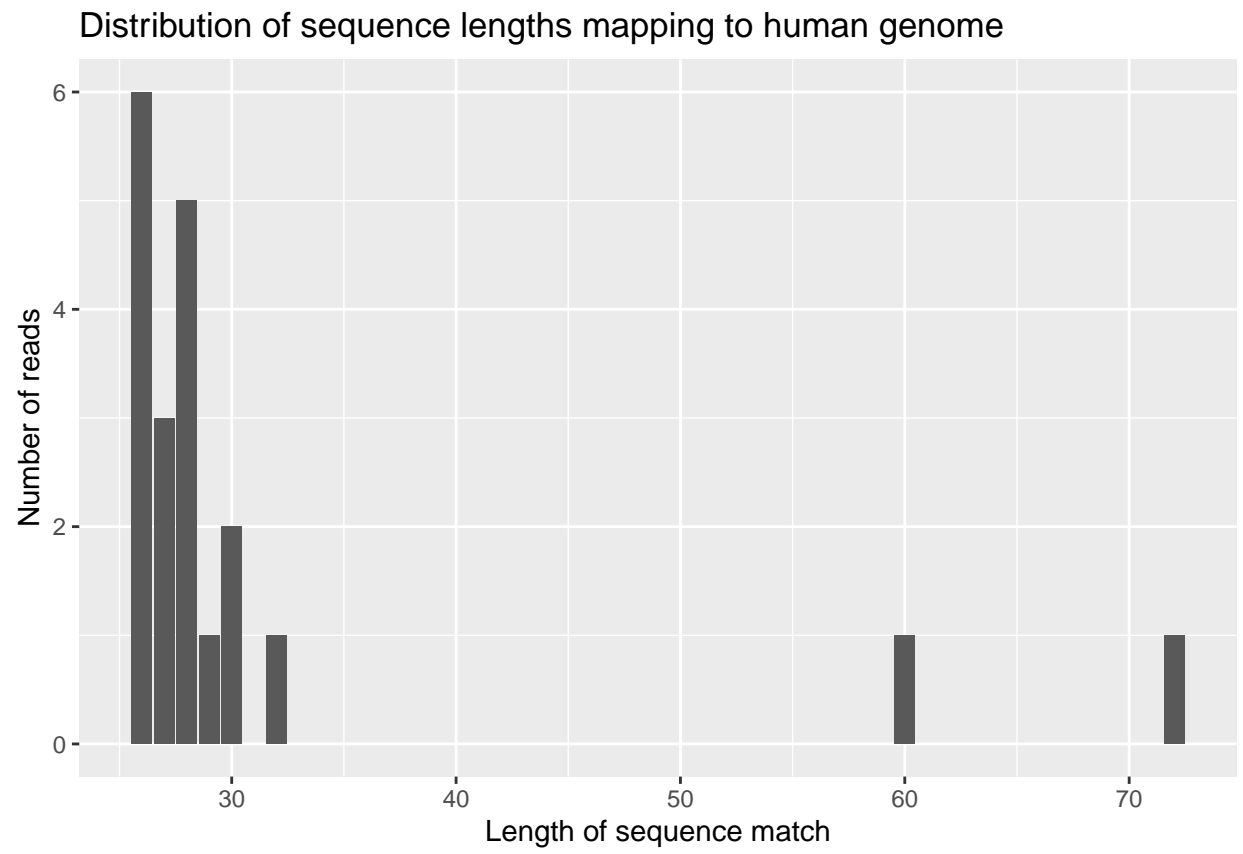
## Key metrics

Total Reads



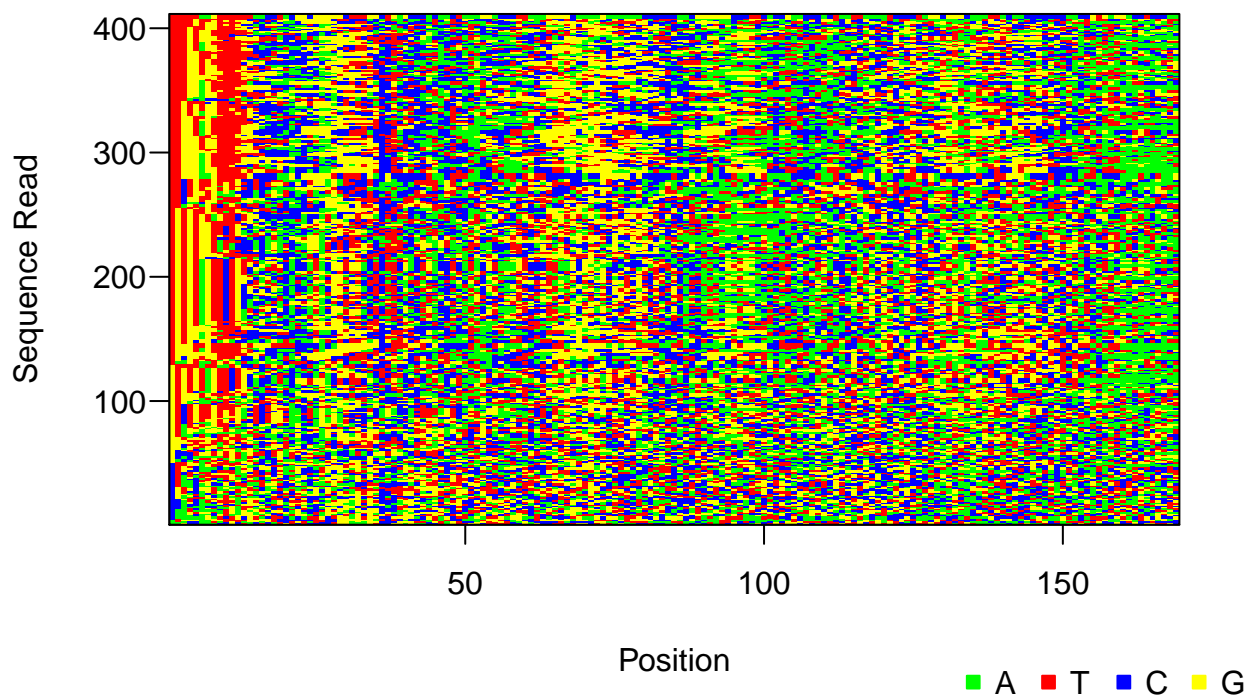
Max Read Proportion



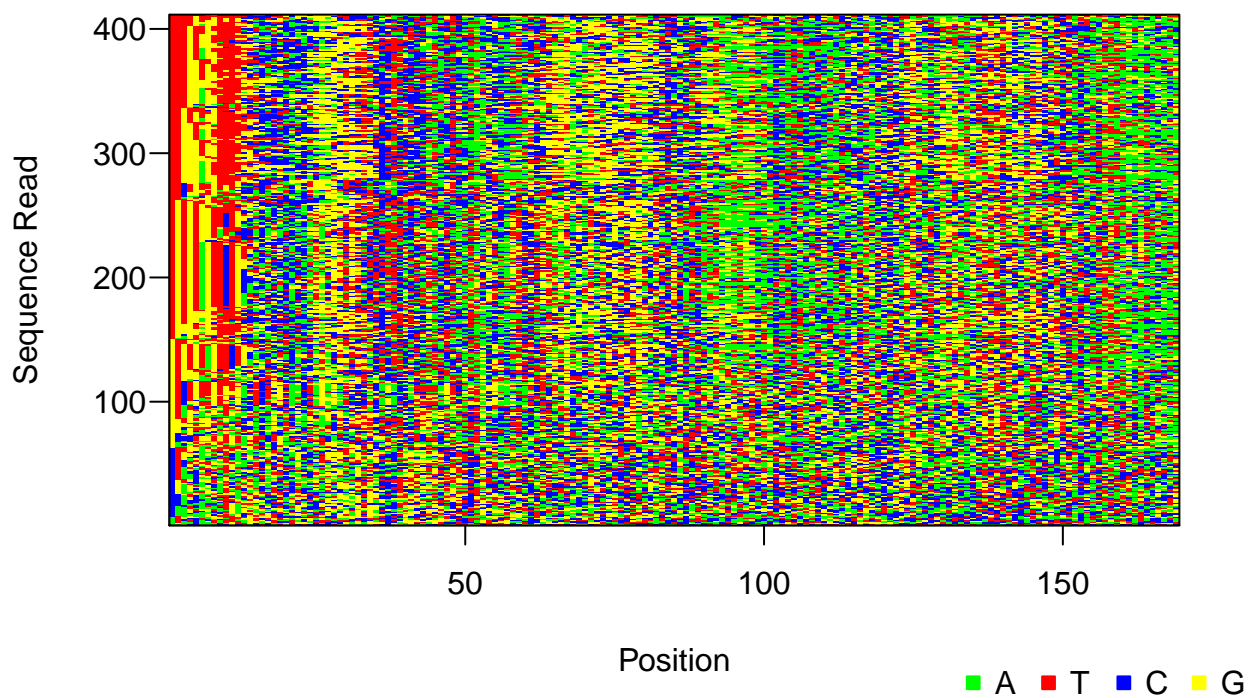


## Visualization of raw FASTQ reads

Random sample of 411 reads, arranged by nucleotide sequence:

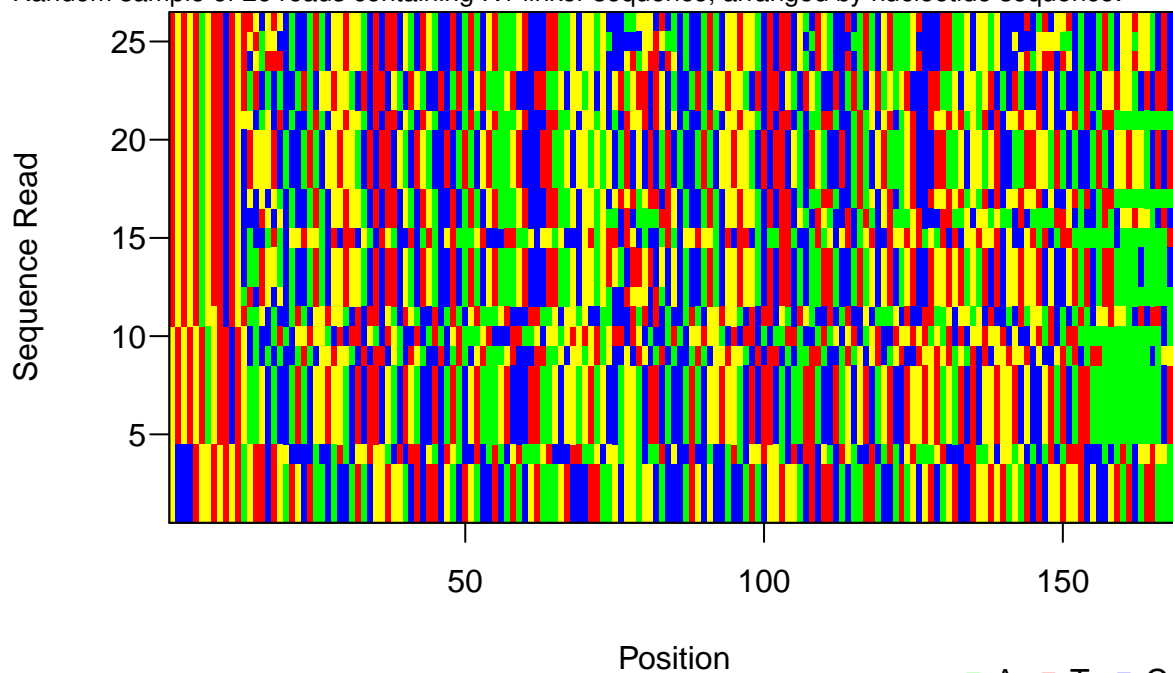


Random sample of 411 distinct reads (i.e. after removing duplicates), arranged by nucleotide sequence:

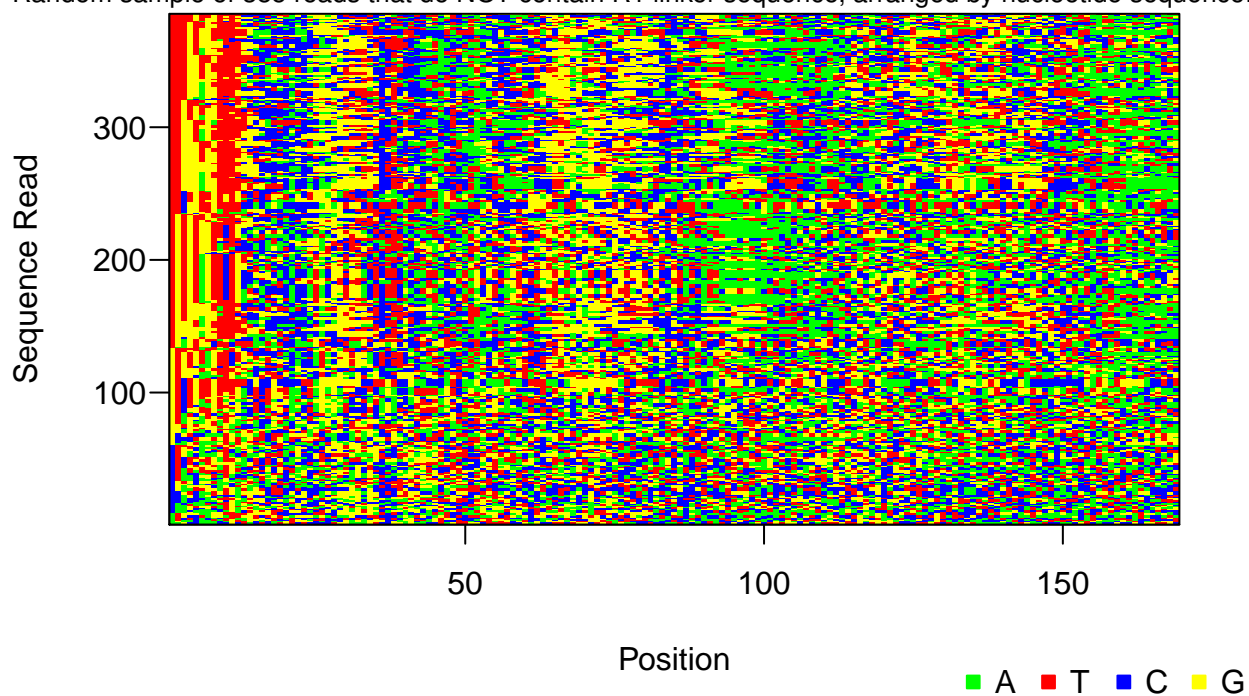


## Visualization of raw FASTQ reads with R1 linker sequences

Random sample of 26 reads containing R1 linker sequence, arranged by nucleotide sequence:



Random sample of 385 reads that do NOT contain R1 linker sequence, arranged by nucleotide sequence:



No primer LTR pairs appeared in more than of total reads.