

## [Record] 180609 (Sat) snakemake --cores 12 filter

**Notebook:** A01工作任務

**Created:** 6/9/2018 2:39 PM

**Updated:** 6/9/2018 6:42 PM

**Author:** clint

**Tags:** @2018 Duke Data+

**URL:** <https://github.com/timflutre/trimmomatic>

```
20 NUM BASES BELOW QUALITY=1 TAG_NAME=XC TAG BARCODED_READ=false HARD_CLIP_BASES=false VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 00:17:31 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
INFO      2018-06-09 00:17:37 TagBamWithReadSequenceExtended Processed 30,000,000 records. Elapsed time: 00:04:55s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:17:38 TagBamWithReadSequenceExtended Processed 1,000,000 records. Elapsed time: 00:00:06s. Time for last 1,000
000:      6s. Last read position: */*
INFO      2018-06-09 00:17:42 TagBamWithReadSequenceExtended Processed 31,000,000 records. Elapsed time: 00:05:01s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:17:43 TagBamWithReadSequenceExtended Processed 2,000,000 records. Elapsed time: 00:00:12s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:01 TagBamWithReadSequenceExtended Processed 32,000,000 records. Elapsed time: 00:05:19s. Time for last 1,000
000:      18s. Last read position: */*
INFO      2018-06-09 00:18:06 TagBamWithReadSequenceExtended Processed 33,000,000 records. Elapsed time: 00:05:25s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:12 TagBamWithReadSequenceExtended Processed 3,000,000 records. Elapsed time: 00:00:40s. Time for last 1,000
000:      28s. Last read position: */*
INFO      2018-06-09 00:18:12 TagBamWithReadSequenceExtended Processed 34,000,000 records. Elapsed time: 00:05:31s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:17 TagBamWithReadSequenceExtended Processed 4,000,000 records. Elapsed time: 00:00:45s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:18 TagBamWithReadSequenceExtended Processed 35,000,000 records. Elapsed time: 00:05:36s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:23 TagBamWithReadSequenceExtended Processed 5,000,000 records. Elapsed time: 00:00:51s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:28 TagBamWithReadSequenceExtended Processed 6,000,000 records. Elapsed time: 00:00:56s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:34 TagBamWithReadSequenceExtended Processed 7,000,000 records. Elapsed time: 00:01:02s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:35 TagBamWithReadSequenceExtended Processed 36,000,000 records. Elapsed time: 00:05:54s. Time for last 1,000
000:      17s. Last read position: */*
INFO      2018-06-09 00:18:41 TagBamWithReadSequenceExtended Processed 37,000,000 records. Elapsed time: 00:05:59s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:46 TagBamWithReadSequenceExtended Processed 38,000,000 records. Elapsed time: 00:06:05s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:18:54 TagBamWithReadSequenceExtended Processed 8,000,000 records. Elapsed time: 00:01:23s. Time for last 1,000
000:      20s. Last read position: */*
INFO      2018-06-09 00:19:00 TagBamWithReadSequenceExtended Processed 9,000,000 records. Elapsed time: 00:01:28s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:19:04 TagBamWithReadSequenceExtended Processed 39,000,000 records. Elapsed time: 00:06:23s. Time for last 1,000
000:      17s. Last read position: */*
INFO      2018-06-09 00:19:10 TagBamWithReadSequenceExtended Processed 40,000,000 records. Elapsed time: 00:06:28s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:19:15 TagBamWithReadSequenceExtended Processed 10,000,000 records. Elapsed time: 00:01:44s. Time for last 1,000
000:      15s. Last read position: */*
INFO      2018-06-09 00:19:16 TagBamWithReadSequenceExtended Processed 41,000,000 records. Elapsed time: 00:06:34s. Time for last 1,000
000:      5s. Last read position: */*
```

```
000:      5s. Last read position: */*
INFO      2018-06-09 00:37:50 TagBamWithReadSequenceExtended Processed 134,000,000 records. Elapsed time: 00:20:18s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:37:55 TagBamWithReadSequenceExtended Processed 135,000,000 records. Elapsed time: 00:20:24s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:38:01 TagBamWithReadSequenceExtended Processed 136,000,000 records. Elapsed time: 00:20:29s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:38:04 TagBamWithReadSequenceExtended Processed 160,000,000 records. Elapsed time: 00:25:22s. Time for last 1,000
000:      17s. Last read position: */*
INFO      2018-06-09 00:38:09 TagBamWithReadSequenceExtended Processed 161,000,000 records. Elapsed time: 00:25:28s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:38:15 TagBamWithReadSequenceExtended Processed 162,000,000 records. Elapsed time: 00:25:34s. Time for last 1,000
000:      5s. Last read position: */*
INFO      2018-06-09 00:38:19 TagBamWithReadSequenceExtended Processed 137,000,000 records. Elapsed time: 00:20:47s. Time for last 1,000
000:      17s. Last read position: */*
INFO      2018-06-09 00:38:24 TagBamWithReadSequenceExtended Processed 138,000,000 records. Elapsed time: 00:20:52s. Time for last 1,000
000:      5s. Last read position: */*
[Sat Jun 09 00:38:27 EDT 2018] org.broadinstitute.dropseqrna.utils.TagBamWithReadSequenceExtended done. Elapsed time: 25.77 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Surface_naph_unaligned.bam.
Finished job 15.
3 of 31 steps (10%) done

rule UMI tags:
  input: data/Surface_naph_BC_tagged_unmapped.bam
  output: data/Surface_naph_BC_UMI_tagged_unmapped.bam, logs/Surface_naph_UMI_barcode.txt
  jobid: 14
  wildcards: sample=Surface_naph

localrule plot_CELL_filtering:
  input: logs/Surface_naph_CELL_barcode.txt
  output: plots/Surface_naph_CELL_dropped.pdf
  jobid: 10
  wildcards: sample=Surface_naph
```





```
Removing temporary output file data/Gland_homeo_unaligned.bam.
Finished job 19.
4 of 31 steps (13%) done
INFO 2018-06-09 00:44:59 TagBamWithReadSequenceExtended Processed 46,000,000 records. Elapsed time: 00:06:31s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:45:04 TagBamWithReadSequenceExtended Processed 47,000,000 records. Elapsed time: 00:06:36s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:45:09 TagBamWithReadSequenceExtended Processed 48,000,000 records. Elapsed time: 00:06:41s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:45:15 TagBamWithReadSequenceExtended Processed 49,000,000 records. Elapsed time: 00:06:46s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:45:31 TagBamWithReadSequenceExtended Processed 50,000,000 records. Elapsed time: 00:07:03s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:45:37 TagBamWithReadSequenceExtended Processed 51,000,000 records. Elapsed time: 00:07:08s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:45:42 TagBamWithReadSequenceExtended Processed 52,000,000 records. Elapsed time: 00:07:13s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:45:47 TagBamWithReadSequenceExtended Processed 53,000,000 records. Elapsed time: 00:07:18s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:46:04 TagBamWithReadSequenceExtended Processed 54,000,000 records. Elapsed time: 00:07:35s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:46:09 TagBamWithReadSequenceExtended Processed 55,000,000 records. Elapsed time: 00:07:40s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:46:14 TagBamWithReadSequenceExtended Processed 56,000,000 records. Elapsed time: 00:07:45s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:46:31 TagBamWithReadSequenceExtended Processed 57,000,000 records. Elapsed time: 00:08:02s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:46:36 TagBamWithReadSequenceExtended Processed 58,000,000 records. Elapsed time: 00:08:07s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:46:41 TagBamWithReadSequenceExtended Processed 59,000,000 records. Elapsed time: 00:08:13s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:46:47 TagBamWithReadSequenceExtended Processed 60,000,000 records. Elapsed time: 00:08:18s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:47:04 TagBamWithReadSequenceExtended Processed 61,000,000 records. Elapsed time: 00:08:35s. Time for last 1,000
000: 17s. Last read position: /*
INFO 2018-06-09 00:47:09 TagBamWithReadSequenceExtended Processed 62,000,000 records. Elapsed time: 00:08:40s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:47:14 TagBamWithReadSequenceExtended Processed 63,000,000 records. Elapsed time: 00:08:45s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:47:19 TagBamWithReadSequenceExtended Processed 64,000,000 records. Elapsed time: 00:08:50s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:47:35 TagBamWithReadSequenceExtended Processed 65,000,000 records. Elapsed time: 00:09:07s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:47:41 TagBamWithReadSequenceExtended Processed 66,000,000 records. Elapsed time: 00:09:12s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:47:46 TagBamWithReadSequenceExtended Processed 67,000,000 records. Elapsed time: 00:09:17s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:48:02 TagBamWithReadSequenceExtended Processed 68,000,000 records. Elapsed time: 00:09:34s. Time for last 1,000
```

```
000: 16s. Last read position: /*
INFO 2018-06-09 00:58:26 TagBamWithReadSequenceExtended Processed 144,000,000 records. Elapsed time: 00:19:57s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:58:31 TagBamWithReadSequenceExtended Processed 145,000,000 records. Elapsed time: 00:20:03s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:58:48 TagBamWithReadSequenceExtended Processed 146,000,000 records. Elapsed time: 00:20:19s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:58:53 TagBamWithReadSequenceExtended Processed 147,000,000 records. Elapsed time: 00:20:24s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:58:58 TagBamWithReadSequenceExtended Processed 148,000,000 records. Elapsed time: 00:20:30s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:59:04 TagBamWithReadSequenceExtended Processed 149,000,000 records. Elapsed time: 00:20:35s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:59:20 TagBamWithReadSequenceExtended Processed 150,000,000 records. Elapsed time: 00:20:52s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:59:25 TagBamWithReadSequenceExtended Processed 151,000,000 records. Elapsed time: 00:20:57s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:59:31 TagBamWithReadSequenceExtended Processed 152,000,000 records. Elapsed time: 00:21:02s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:59:36 TagBamWithReadSequenceExtended Processed 153,000,000 records. Elapsed time: 00:21:07s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 00:59:53 TagBamWithReadSequenceExtended Processed 154,000,000 records. Elapsed time: 00:21:24s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 00:59:58 TagBamWithReadSequenceExtended Processed 155,000,000 records. Elapsed time: 00:21:29s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 01:00:03 TagBamWithReadSequenceExtended Processed 156,000,000 records. Elapsed time: 00:21:34s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 01:00:19 TagBamWithReadSequenceExtended Processed 157,000,000 records. Elapsed time: 00:21:51s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 01:00:24 TagBamWithReadSequenceExtended Processed 158,000,000 records. Elapsed time: 00:21:56s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 01:00:30 TagBamWithReadSequenceExtended Processed 159,000,000 records. Elapsed time: 00:22:01s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 01:00:35 TagBamWithReadSequenceExtended Processed 160,000,000 records. Elapsed time: 00:22:06s. Time for last 1,000
000: 5s. Last read position: /*
INFO 2018-06-09 01:00:51 TagBamWithReadSequenceExtended Processed 161,000,000 records. Elapsed time: 00:22:23s. Time for last 1,000
000: 16s. Last read position: /*
INFO 2018-06-09 01:00:56 TagBamWithReadSequenceExtended Processed 162,000,000 records. Elapsed time: 00:22:28s. Time for last 1,000
000: 5s. Last read position: /*
```

```
[Sat Jun 09 01:01:01 EDT 2018] org.broadinstitute.dropseqrna.utils.TagBamWithReadSequenceExtended done. Elapsed time: 22.55 minutes.
Runtime.totalMemory()=1485307904
```

```
Removing temporary output file data/Surface_naph_BC_tagged_unmapped.bam.
```

```
Finished job 14.
```

```
5 of 31 steps (16%) done
```

```
Shutting down, this might take some time.
```

```
Exiting because a job execution failed. Look above for error message
```

```
Complete log: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/log/2018-06-08T231104.591490.snakemake.log
```

```
(env-py3) clintko@phl-biostats-ub1: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$
```

```
[Sat Jun 09 01:01:01 EDT 2018] org.broadinstitute.dropseqrna.utils.TagBamWithReadSequenceExtended done. Elapsed time: 22.55 minutes.
```

```
Runtime.totalMemory()=1485307904
```

```
Removing temporary output file data/Surface_naph_BC_tagged_unmapped.bam.
```

```
Finished job 14.
```

```
5 of 31 steps (16%) done
```

```
Shutting down, this might take some time.
```

```
Exiting because a job execution failed. Look above for error message
```

```
Complete log: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/log/2018-06-08T231104.591490.snakemake.log
```

```
(env-py3) clintko@phl-biostats-ub1: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$
```

<pre> localrule plot_CELL_filtering:   input: logs/Surface_naph_CELL_barcode.txt   output: plots/Surface_naph_CELL_dropped.pdf   jobid: 10   wilddcards: sample=Surface_naph  + java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar Ingenuity dSequenceExtended SUMMARY=logs/Surface_naph_UMI_barcode.txt BASE_RANGE=17-24 BASE_QUALITY=20 BARCODED_READ=1 DISCARD_READ=true TAG_NAME=XM NUM BASES_BELOW_QUALITY=1 INPUT=/data/Surface_naph_BC_UMI_tagged_unmapped.bam OUTPUT=/data/Surface_naph_BC_UMI_tagged_unmapped.bam [Sat Jun 09 00:38:28 EDT 2018] org.broadinstitute.dropseqrn.uti OUTPUT=/data/Surface_naph_BC_UMI_tagged_unmapped.bam SUMMARY=log e BASE_QUALITY=20 NUM_BASES_BELOW_QUALITY=1 TAG_NAME=XM TAG_B TRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 C [Sat Jun 09 00:38:28 EDT 2018] Executing as clintko@phl-biostats h11-Aubuntu0 17.10.1-h11: Picard version: 1.13(7hed8f4_151300803 Error in library(ggpubr) : there is no package called 'ggpubr' Execution halted Error in rule plot_CELL_filtering:   jobid: 10   output: plots/Surface_naph_CELL_dropped.pdf </pre>	<pre> 227 rule plot_CELL_filtering: 228     input: 229         'logs/{sample}_CELL_barcode.txt' 230 231     params: 232         min_quality=config['FILTER']['cell-barcode']['min-quality'], 233         num_below_quality=config['FILTER']['cell-barcode']['num-below-quality'] 234 235     conda: './envs/plots.yaml' 236 237     output: 238         pdf='plots/{sample}_CELL_dropped.pdf' 239 240     script: 241         '../scripts/plot_cell_drop.R' </pre>
<pre> localrule plot_CELL_filtering:   input: logs/Surface_naph_CELL_barcode.txt   output: plots/Surface_naph_CELL_dropped.pdf   jobid: 10   wilddcards: sample=Surface_naph  + java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp - dSequenceExtended SUMMARY=logs/Surface_naph_UMI_barcode.txt BASE ASES_BELOW_QUALITY=1 INPUT=/data/Surface_naph_BC_UMI_tagged_unmapped. [Sat Jun 09 00:38:28 EDT 2018] org.broadinstitute.dropseqrn.uti OUTPUT=/data/Surface_naph_BC_UMI_tagged_unmapped.bam SUMMARY=log e BASE_QUALITY=20 NUM_BASES_BELOW_QUALITY=1 TAG_NAME=XM TAG_B TRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 C [Sat Jun 09 00:38:28 EDT 2018] Executing as clintko@phl-biostats h11-Aubuntu0 17.10.1-h11: Picard version: 1.13(7hed8f4_151300803 Error in library(ggpubr) : there is no package called 'ggpubr' Execution halted Error in rule plot_CELL_filtering:   jobid: 10   output: plots/Surface_naph_CELL_dropped.pdf </pre>	
<ul style="list-style-type: none"> <li>merge_counts.R</li> <li>merge_counts_single.R</li> <li>plot_base_freq.R</li> <li>plot_BC_drop.R</li> <li>plot_cell_drop.R</li> <li>plot_knee_plot.R</li> <li>plot_polyA_trim.R</li> <li>plot_rna_metrics.R</li> <li>plot_species_plot.R</li> <li>plot_start_trim.R</li> <li>plot_umi_distribution.R</li> <li>plot_umi_drop.R</li> <li>plot_yield.R</li> </ul>	<pre> library(ggplot2) library(ggpubr) data = read.table(file = snakemake@input[[1]], header=T, stringsAsFactors=F) data\$presence=rep('kept', nrow(data)) ids = which(snakemake@params\$num_below_quality &lt; data\$num_failed_bases) data\$presence[ids] = 'tagged' tagged = sum(data\$num_barcodes[ids]) p = ggplot(data, aes(x=num_failed_bases, y=num_barcodes, fill = presence)) p = p + geom_bar(stat = 'identity') p = p + ggtitle(paste('Number of CELL barcode bases under', snakemake@params\$min_quality,'quality in\n', s p = p + labs(x='Num of failed bases', y='Counts') p = p + theme_pubr(legend = 'bottom') p = p + scale_x_continuous(breaks = c(data\$num_failed_bases), labels = factor(data\$num_failed_bases)) p = p + geom_vline(aes(xintercept = snakemake@params\$num_below_quality + 0.5), color = 'red') p = p + scale_y_continuous(labels = scales::scientific) ggsave(plot = p, snakemake@output\$pdf, height = 4, width = 6) </pre>

rerun again

<pre> (env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe\$ snakemake --cores 12 filter Building DAG of jobs... Using shell: /bin/bash Provided cores: 12 Rules claiming more threads will be scaled down. Job counts:   count  jobs     1     UMI tags     2     count reads     1     filter     2     filter_tags     1     multiqc trimmomatic     1     plot_BC_drop     2     plot_CELL_filtering     2     plot_UMI_filtering     2     plot_barcode_start_trim     2     plot_polyA_trim     2     polyA_trim     2     reads_after_trimming     2     sam_to_fastq     2     start_trim     2     trim_single     26 </pre>	
--	--

```

localrule plot_CELL_filtering:
  input: logs/Gland_homeo_CELL_barcode.txt
  output: plots/Gland_homeo_CELL_dropped.pdf
  jobid: 2
  wildcards: sample=Gland_homeo

rule filter_tags:
  input: data/Surface_naph_BC_UMI_tagged_unmapped.bam
  output: data/Surface_naph_tags_filtered_unmapped.bam
  jobid: 28
  wildcards: sample=Surface_naph

rule UMI_tags:
  input: data/Gland_homeo_BC_tagged_unmapped.bam
  output: data/Gland_homeo_BC_UMI_tagged_unmapped.bam, logs/Gland_homeo_UMI_barcode.txt
  jobid: 18
  wildcards: sample=Gland_homeo

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar FilterBAM TAG
REJECT=XQ INPUT=data/Surface_naph_BC_UMI_tagged_unmapped.bam OUTPUT=data/Surface_naph_tags_filtered_unmapped.bam

```

```

localrule plot_UMI_filtering:
  input: logs/Surface_naph_UMI_barcode.txt
  output: plots/Surface_naph_UMI_dropped.pdf
  jobid: 3
  wildcards: sample=Surface_naph

localrule plot_CELL_filtering:
  input: logs/Surface_naph_CELL_barcode.txt
  output: plots/Surface_naph_CELL_dropped.pdf
  jobid: 11
  wildcards: sample=Surface_naph

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar TagBamWithReadSequenceExtended SUMMARY=logs/Gland_homeo_UMI_barcode.txt BASE_RANGE=17-24 BASE_QUALITY=20 BARCODED_READ=1 DISCARD_READ=true TAG_NAME=XM NUM_BAS
SES_BELOW_QUALITY=1 INPUT=data/Gland_homeo_BC_tagged_unmapped.bam OUTPUT=data/Gland_homeo_BC_UMI_tagged_unmapped.bam
[Sat Jun 09 14:54:33 EDT 2018] org.broadinstitute.dropseqrna.utils.FilterBAM INPUT=data/Surface_naph_BC_UMI_tagged_unmapped.bam OUTPUT=data/Sur
ace_naph_tags_filtered_unmapped.bam TAG_REJECT=[XQ] FILTER_PCR_DUPE=false RETAIN_ONLY_PRIMARY_READS=false DROP_REJECTED_REF=false VERBOSITY=
INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT
T_SECRETS=client_secrets.json
[Sat Jun 09 14:54:33 EDT 2018] org.broadinstitute.dropseqrna.utils.TagBamWithReadSequenceExtended INPUT=data/Gland_homeo_BC_tagged_unmapped.bam
OUTPUT=data/Gland_homeo_BC_UMI_tagged_unmapped.bam SUMMARY=logs/Gland_homeo_UMI_barcode.txt BASE_RANGE=17-24 BARCODED_READ=1 DISCARD_READ=true B
ASE_QUALITY=20 NUM_BASSES_BELOW_QUALITY=1 TAG_NAME=XM TAG_BARCODED_READ=false HARD_CLIP_BASSES=false VERBOSITY=INFO QUIET=false VALIDATION_STR
NGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 14:54:33 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171
b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
[Sat Jun 09 14:54:33 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171
b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
Loading required package: magrittr
Loading required package: magrittr
Loading required package: magrittr
Finished job 2.
1 of 26 steps (4%) done
Finished job 11.
2 of 26 steps (8%) done
Finished job 3.
3 of 26 steps (12%) done

```

```

Finished job 2.
1 of 26 steps (4%) done
Finished job 11.
2 of 26 steps (8%) done
Finished job 3.
3 of 26 steps (12%) done
INFO 2018-06-09 14:54:43 TagBamWithReadSequenceExtended Processed 1,000,000 records. Elapsed time: 00:00:09s. Time for last 1,000,000: 9s. Last read position: */*
INFO 2018-06-09 14:54:44 FilterBAM Processed 1,000,000 records. Elapsed time: 00:00:11s. Time for last 1,000,000: 11s. Last read position: */*
INFO 2018-06-09 14:54:48 TagBamWithReadSequenceExtended Processed 2,000,000 records. Elapsed time: 00:00:15s. Time for last 1,000,000: 15s. Last read position: */*
INFO 2018-06-09 14:54:51 FilterBAM Processed 2,000,000 records. Elapsed time: 00:00:18s. Time for last 1,000,000: 6s. Last read position: */*
INFO 2018-06-09 14:54:53 TagBamWithReadSequenceExtended Processed 3,000,000 records. Elapsed time: 00:00:20s. Time for last 1,000,000: 20s. Last read position: */*
INFO 2018-06-09 14:54:58 FilterBAM Processed 3,000,000 records. Elapsed time: 00:00:24s. Time for last 1,000,000: 6s. Last read position: */*
INFO 2018-06-09 14:54:58 TagBamWithReadSequenceExtended Processed 4,000,000 records. Elapsed time: 00:00:25s. Time for last 1,000,000: 25s. Last read position: */*
INFO 2018-06-09 14:55:03 TagBamWithReadSequenceExtended Processed 5,000,000 records. Elapsed time: 00:00:30s. Time for last 1,000,000: 30s. Last read position: */*
INFO 2018-06-09 14:55:04 FilterBAM Processed 4,000,000 records. Elapsed time: 00:00:31s. Time for last 1,000,000: 6s. Last read position: */*

```

```

000: 5s. Last read position: */*
INFO 2018-06-09 15:08:59 TagBamWithReadSequenceExtended Processed 105,000,000 records. Elapsed time: 00:14:26s. Time for last 1,000,000: 19s. Last read position: */*
000: 5s. Last read position: */*
INFO 2018-06-09 15:09:05 FilterBAM Processed 77,000,000 records. Elapsed time: 00:14:32s. Time for last 1,000,000: 19s. Last read position: */*
t read position: */*
INFO 2018-06-09 15:09:11 FilterBAM Processed 78,000,000 records. Elapsed time: 00:14:38s. Time for last 1,000,000: 6s. Last read position: */*
t read position: */*
INFO 2018-06-09 15:09:15 TagBamWithReadSequenceExtended Processed 106,000,000 records. Elapsed time: 00:14:42s. Time for last 1,000,000: 19s. Last read position: */*
000: 16s. Last read position: */*
INFO 2018-06-09 15:09:21 TagBamWithReadSequenceExtended Processed 107,000,000 records. Elapsed time: 00:14:47s. Time for last 1,000,000: 19s. Last read position: */*
000: 5s. Last read position: */*
INFO 2018-06-09 15:09:26 TagBamWithReadSequenceExtended Processed 108,000,000 records. Elapsed time: 00:14:52s. Time for last 1,000,000: 19s. Last read position: */*
000: 5s. Last read position: */*
INFO 2018-06-09 15:09:30 FilterBAM Processed 79,000,000 records. Elapsed time: 00:14:57s. Time for last 1,000,000: 19s. Last read position: */*
t read position: */*
INFO 2018-06-09 15:09:31 TagBamWithReadSequenceExtended Processed 109,000,000 records. Elapsed time: 00:14:57s. Time for last 1,000,000: 19s. Last read position: */*
000: 5s. Last read position: */*
INFO 2018-06-09 15:09:37 FilterBAM Processed 80,000,000 records. Elapsed time: 00:15:04s. Time for last 1,000,000: 6s. Last read position: */*
t read position: */*
INFO 2018-06-09 15:09:43 FilterBAM Processed 81,000,000 records. Elapsed time: 00:15:10s. Time for last 1,000,000: 6s. Last read position: */*
t read position: */*
INFO 2018-06-09 15:09:48 TagBamWithReadSequenceExtended Processed 110,000,000 records. Elapsed time: 00:15:14s. Time for last 1,000,000: 19s. Last read position: */*
000: 16s. Last read position: */*
INFO 2018-06-09 15:09:53 TagBamWithReadSequenceExtended Processed 111,000,000 records. Elapsed time: 00:15:19s. Time for last 1,000,000: 19s. Last read position: */*
000: 4s. Last read position: */*
[Sat Jun 09 15:09:54 EDT 2018] org.broadinstitute.dropseqrna.utils.FilterBAM done. Elapsed time: 15.35 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Surface_naph_BC_UMI_tagged_unmapped.bam.
Finished job 28.
4 of 26 steps (15%) done

```

## error found: samtools: command not found

```

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar TrimStarting
sequence OUTPUT_SUMMARY=logs/Surface_naph_start_trim.txt SEQUENCE=AATGATACGGCGACCGAGATCTACACGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC MISMATCHES=
NUM_BASES=6 INPUT=data/Surface_naph_tags_filtered_unmapped.bam OUTPUT=data/Surface_naph_tags_start_filtered_unmapped.bam
/bin/bash: samtools: command not found
Error in rule count_reads:
  jobid: 21
  output: logs/Surface_naph_reads_left.txt

RuleException:
CalledProcessError in line 102 of /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk:
Command 'set -euo pipefail; samtools view data/Surface_naph_tags_filtered_unmapped.bam | wc -l > logs/Surface_naph_reads_left.txt' returned
on-zero exit status 127.
File "/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk", line 102, in __rule_count_reads
File "/usr/lib/python3.6/concurrent/futures/thread.py", line 56, in run
Removing output files of failed job count_reads since they might be corrupted:
logs/Surface_naph_reads_left.txt

```

fixed the problems

```

(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ alias samtools='samtools_0.1.18'
(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ samtools

Program: samtools (Tools for alignments in the SAM format)
Version: 0.1.18 (r982:295)

Usage: samtools <command> [options]

Command: view      SAM<->BAM conversion
          sort      sort alignment file
          mpileup    multi-way pileup
          depth      compute the depth
          faidx      index/extract FASTA
          index      index alignment
          idxstats   BAM index stats (r595 or later)
          fixmate     fix mate information
          flagstat    simple stats
          calmd       recalculate MD/NM tags and '=' bases
          merge       merge sorted alignments
          rmdup       remove PCR duplicates
          reheader    replace BAM header
          cat         concatenate BAMs
          targetcut   cut fosmid regions (for fosmid pool only)
          phase       phase heterozygotes

```

the problem still appear!!!

the alias method does not work

I have turned off the terminal and open another one, try it again, the error still occurs

Therefore, I have no chose but hard code the samtools in the snakefile



```

rule count_reads:
    input:
        'data/{sample}_tags_filtered_unmapped.bam'
    output:
        'logs/{sample}_reads_left.txt'
    conda: ' ../envs/santools.yaml'
    shell:
        """santools_0.1.18 view {input} | wc -l > {output}"""

```

In this time, it does not complaint about samtools command not found again

```

(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ snakemake --cores 12 filter
Building DAG of jobs...
Using shell: /bin/bash
Provided cores: 12
Rules claiming more threads will be scaled down.
Job counts:
   count  jobs
    2     count_reads
    1     filter
    1     filter_tags
    1     multiqc_trimmomatic
    1     plot_BC_drop
    2     plot_barcode_start_trim
    2     plot_polyA_trim
    2     polyA_trim
    2     reads_after_trimming
    2     sam_to_fastq
    2     start_trim
    2     trim_single
    20
localrule count_reads:
  input: data/Surface_naph_tags_filtered_unmapped.bam
  output: logs/Surface_naph_reads_left.txt
  jobid: 23
  wildcards: sample=Surface_naph

```

```

(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ snakemake --cores 12 filter
Building DAG of jobs...
Using shell: /bin/bash
Provided cores: 12
Rules claiming more threads will be scaled down.
Job counts:
   count  jobs
    2     count_reads
    1     filter
    1     filter_tags
    1     multiqc_trimmomatic
    1     plot_BC_drop
    2     plot_barcode_start_trim
    2     plot_polyA_trim
    2     polyA_trim
    2     reads_after_trimming
    2     sam_to_fastq
    2     start_trim
    2     trim_single
    20
localrule count_reads:
  input: data/Surface_naph_tags_filtered_unmapped.bam
  output: logs/Surface_naph_reads_left.txt
  jobid: 23
  wildcards: sample=Surface_naph

rule start_trim:
  input: data/Surface_naph_tags_filtered_unmapped.bam
  output: data/Surface_naph_tags_start_filtered_unmapped.bam, logs/Surface_naph_start_trim.txt
  jobid: 15
  wildcards: sample=Surface_naph

```

```
rule filter_tags:
  input: data/Gland_homeo_BC_UMI_tagged_unmapped.bam
  output: data/Gland_homeo_tags_filtered_unmapped.bam
  jobid: 30
  wildcards: sample=Gland_homeo

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar TrimStartingSequence OUTPUT=logs/Surface_naph_start_trim.txt SEQUENCE=AATGATACGGCGACCACCGAGATCTACAGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC MISMATCHES=1 NUM_BASES=6 INPUT=data/Surface_naph_tags_filtered_unmapped.bam OUTPUT=data/Surface_naph_tags_start_filtered_unmapped.bam
+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar FilterBAM TAG_REJECT=XQ INPUT=data/Gland_homeo_BC_UMI_tagged_unmapped.bam OUTPUT=data/Gland_homeo_tags_filtered_unmapped.bam
[Sat Jun 09 15:47:40 EDT 2018] org.broadinstitute.dropseqrna.utils.FilterBAM INPUT=data/Gland_homeo_BC_UMI_tagged_unmapped.bam OUTPUT=data/Gland_homeo_tags_filtered_unmapped.bam TAG_REJECT=[XQ] FILTER_PCR_DUPES=false RETAIN_ONLY_PRIMARY_READS=false DROP_REJECTED_REF=false VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 15:47:40 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
[Sat Jun 09 15:47:40 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.TrimStartingSequence INPUT=data/Surface_naph_tags_filtered_unmapped.bam OUTPUT=data/Surface_naph_tags_start_filtered_unmapped.bam OUTPUT=logs/Surface_naph_start_trim.txt SEQUENCE=AATGATACGGCGACCACCGAGATCTACAGCCTGTCCGCGGAAGCAGTGGTATCAACGCAGAGTAC MISMATCHES=1 NUM_BASES=6 TRIM_TAG=ZS VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 15:47:40 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
INFO 2018-06-09 15:47:47 FilterBAM Processed 1,000,000 records. Elapsed time: 00:00:07s. Time for last 1,000,000: 7s. Last read position: /*/
INFO 2018-06-09 15:47:50 TrimStartingSequence Processed 1,000,000 records. Elapsed time: 00:00:10s. Time for last 1,000,000: 10s. Last read position: /*/
INFO 2018-06-09 15:47:54 FilterBAM Processed 2,000,000 records. Elapsed time: 00:00:13s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:48:00 TrimStartingSequence Processed 2,000,000 records. Elapsed time: 00:00:19s. Time for last 1,000,000: 19s. Last read position: /*/
INFO 2018-06-09 15:48:00 FilterBAM Processed 3,000,000 records. Elapsed time: 00:00:19s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:48:06 FilterBAM Processed 4,000,000 records. Elapsed time: 00:00:25s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:48:09 TrimStartingSequence Processed 3,000,000 records. Elapsed time: 00:00:28s. Time for last 1,000,000: 28s. Last read position: /*/
INFO 2018-06-09 15:48:12 FilterBAM Processed 5,000,000 records. Elapsed time: 00:00:31s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:49:02 TrimStartingSequence Processed 4,000,000 records. Elapsed time: 00:01:21s. Time for last 1,000,000: 51s. Last read position: /*/
INFO 2018-06-09 15:49:04 FilterBAM Processed 6,000,000 records. Elapsed time: 00:01:24s. Time for last 1,000,000: 52s. Last read position: /*/
Finished job 23.
1 of 20 steps (5%) done
```

```
t read position: /*/
INFO 2018-06-09 15:48:06 FilterBAM Processed 4,000,000 records. Elapsed time: 00:00:25s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:48:09 TrimStartingSequence Processed 3,000,000 records. Elapsed time: 00:00:28s. Time for last 1,000,000: 28s. Last read position: /*/
INFO 2018-06-09 15:48:12 FilterBAM Processed 5,000,000 records. Elapsed time: 00:00:31s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:49:02 TrimStartingSequence Processed 4,000,000 records. Elapsed time: 00:01:21s. Time for last 1,000,000: 51s. Last read position: /*/
INFO 2018-06-09 15:49:04 FilterBAM Processed 6,000,000 records. Elapsed time: 00:01:24s. Time for last 1,000,000: 52s. Last read position: /*/
Finished job 23.
1 of 20 steps (5%) done
INFO 2018-06-09 15:49:10 FilterBAM Processed 7,000,000 records. Elapsed time: 00:01:30s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:49:11 TrimStartingSequence Processed 5,000,000 records. Elapsed time: 00:01:30s. Time for last 1,000,000: 30s. Last read position: /*/
INFO 2018-06-09 15:49:20 TrimStartingSequence Processed 6,000,000 records. Elapsed time: 00:01:39s. Time for last 1,000,000: 39s. Last read position: /*/
INFO 2018-06-09 15:49:28 FilterBAM Processed 8,000,000 records. Elapsed time: 00:01:47s. Time for last 1,000,000: 17s. Last read position: /*/
INFO 2018-06-09 15:49:34 FilterBAM Processed 9,000,000 records. Elapsed time: 00:01:53s. Time for last 1,000,000: 5s. Last read position: /*/
INFO 2018-06-09 15:49:40 FilterBAM Processed 10,000,000 records. Elapsed time: 00:01:59s. Time for last 1,000,000: 5s. Last read position: /*/
INFO 2018-06-09 15:49:45 FilterBAM Processed 11,000,000 records. Elapsed time: 00:02:05s. Time for last 1,000,000: 5s. Last read position: /*/
INFO 2018-06-09 15:49:49 TrimStartingSequence Processed 7,000,000 records. Elapsed time: 00:02:09s. Time for last 1,000,000: 29s. Last read position: /*/
INFO 2018-06-09 15:49:58 TrimStartingSequence Processed 8,000,000 records. Elapsed time: 00:02:18s. Time for last 1,000,000: 18s. Last read position: /*/
INFO 2018-06-09 15:50:06 FilterBAM Processed 12,000,000 records. Elapsed time: 00:02:25s. Time for last 1,000,000: 20s. Last read position: /*/
INFO 2018-06-09 15:50:12 FilterBAM Processed 13,000,000 records. Elapsed time: 00:02:31s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:50:21 TrimStartingSequence Processed 9,000,000 records. Elapsed time: 00:02:40s. Time for last 1,000,000: 40s. Last read position: /*/
INFO 2018-06-09 15:50:30 TrimStartingSequence Processed 10,000,000 records. Elapsed time: 00:02:50s. Time for last 1,000,000: 50s. Last read position: /*/
INFO 2018-06-09 15:50:30 FilterBAM Processed 14,000,000 records. Elapsed time: 00:02:50s. Time for last 1,000,000: 18s. Last read position: /*/
INFO 2018-06-09 15:50:36 FilterBAM Processed 15,000,000 records. Elapsed time: 00:02:56s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:50:42 FilterBAM Processed 16,000,000 records. Elapsed time: 00:03:02s. Time for last 1,000,000: 6s. Last read position: /*/
INFO 2018-06-09 15:50:51 TrimStartingSequence Processed 11,000,000 records. Elapsed time: 00:03:10s. Time for last 1,000,000: 10s. Last read position: /*/
INFO 2018-06-09 15:51:00 TrimStartingSequence Processed 12,000,000 records. Elapsed time: 00:03:19s. Time for last 1,000,000: 19s. Last read position: /*/
```



```

t read position: /*/
INFO 2018-06-09 16:02:28 FilterBAM Processed 83,000,000 records. Elapsed time: 00:14:48s. Time for last 1,000,000: 6s. La
t read position: /*/
INFO 2018-06-09 16:02:39 TrimStartingSequence Processed 58,000,000 records. Elapsed time: 00:14:58s. Time for last 1,000,000:
1s. Last read position: /*/
INFO 2018-06-09 16:02:48 FilterBAM Processed 84,000,000 records. Elapsed time: 00:15:07s. Time for last 1,000,000: 19s. La
t read position: /*/
INFO 2018-06-09 16:02:48 TrimStartingSequence Processed 59,000,000 records. Elapsed time: 00:15:08s. Time for last 1,000,000:
9s. Last read position: /*/
INFO 2018-06-09 16:02:54 FilterBAM Processed 85,000,000 records. Elapsed time: 00:15:13s. Time for last 1,000,000: 6s. La
t read position: /*/
INFO 2018-06-09 16:03:00 FilterBAM Processed 86,000,000 records. Elapsed time: 00:15:19s. Time for last 1,000,000: 6s. La
t read position: /*/
INFO 2018-06-09 16:03:09 TrimStartingSequence Processed 60,000,000 records. Elapsed time: 00:15:29s. Time for last 1,000,000:
1s. Last read position: /*/
INFO 2018-06-09 16:03:19 TrimStartingSequence Processed 61,000,000 records. Elapsed time: 00:15:38s. Time for last 1,000,000:
9s. Last read position: /*/
INFO 2018-06-09 16:03:19 FilterBAM Processed 87,000,000 records. Elapsed time: 00:15:38s. Time for last 1,000,000: 18s. La
t read position: /*/
INFO 2018-06-09 16:03:25 FilterBAM Processed 88,000,000 records. Elapsed time: 00:15:44s. Time for last 1,000,000: 6s. La
t read position: /*/
[Sat Jun 09 16:03:39 EDT 2018] org.broadinstitute.dropseqrna.utils.FilterBAM done. Elapsed time: 15.98 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Gland_homeo_BC_UMI_tagged_unmapped.bam.
Finished job 30.
2 of 20 steps (10%) done

rule start_trim:
  input: data/Gland_homeo_tags_filtered_unmapped.bam
  output: data/Gland_homeo_tags_start_filtered_unmapped.bam, logs/Gland_homeo_start_trim.txt
  jobid: 25
  wildcards: sample=Gland_homeo

localrule count_reads:
  input: data/Gland_homeo_tags_filtered_unmapped.bam
  output: logs/Gland_homeo_reads_left.txt
  jobid: 22
  wildcards: sample=Gland_homeo

```

```

localrule count_reads:
  input: data/Gland_homeo_tags_filtered_unmapped.bam
  output: logs/Gland_homeo_reads_left.txt
  jobid: 22
  wildcards: sample=Gland_homeo

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar TrimStartingSequence OUTPUT_SUMMARY=logs/Gland_homeo_start_trim.txt SEQUENCE=AATGATACGGCGACCGAGATCTACACGCTGTCCGCGAAGCAGTGGTATCAACGCAGAGTAC MISMATCHES=1 NUM_BASES=6 INPUT=data/Gland_homeo_tags_filtered_unmapped.bam OUTPUT=data/Gland_homeo_tags_start_filtered_unmapped.bam [Sat Jun 09 16:03:40 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.TrimStartingSequence INPUT=data/Gland_homeo_tags_filtered_unmapped.bam OUTPUT=data/Gland_homeo_tags_start_filtered_unmapped.bam OUTPUT_SUMMARY=logs/Gland_homeo_start_trim.txt SEQUENCE=AATGATACGGCGACCGAGATCTACACGCTGTCCGCGAAGCAGTGGTATCAACGCAGAGTAC MISMATCHES=1 NUM_BASES=6 TRIM_TAG=ZS VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json [Sat Jun 09 16:03:40 EDT 2018] Executing as clintko@phl-biosTats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
INFO 2018-06-09 16:03:45 TrimStartingSequence Processed 62,000,000 records. Elapsed time: 00:16:04s. Time for last 1,000,000: 2
6s. Last read position: /*/
INFO 2018-06-09 16:03:50 TrimStartingSequence Processed 1,000,000 records. Elapsed time: 00:00:09s. Time for last 1,000,000:
9s. Last read position: /*/
INFO 2018-06-09 16:03:58 TrimStartingSequence Processed 2,000,000 records. Elapsed time: 00:00:18s. Time for last 1,000,000:
8s. Last read position: /*/
INFO 2018-06-09 16:04:03 TrimStartingSequence Processed 63,000,000 records. Elapsed time: 00:16:22s. Time for last 1,000,000: 1
7s. Last read position: /*/
INFO 2018-06-09 16:04:07 TrimStartingSequence Processed 3,000,000 records. Elapsed time: 00:00:26s. Time for last 1,000,000:
8s. Last read position: /*/
INFO 2018-06-09 16:04:17 TrimStartingSequence Number of reads trimmed: 3012437 total reads: 63605917
[Sat Jun 09 16:04:18 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.TrimStartingSequence done. Elapsed time: 16.63 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Surface_naph_tags_filtered_unmapped.bam.
Finished job 15.
3 of 20 steps (15%) done

```

```

localrule plot_barcode_start_trim:
  input: logs/Surface_naph_start_trim.txt
  output: plots/Surface_naph_start_trim.pdf
  jobid: 3
  wildcards: sample=Surface_naph

rule polyA_trim:
  input: data/Surface_naph_tags_start_filtered_unmapped.bam
  output: data/Surface_naph_trimmed_unmapped.bam, logs/Surface_naph_polyA_trim.txt
  jobid: 14
  wildcards: sample=Surface_naph

Loading required package: magrittr
+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq_Lung/tmp -jar /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/jar/dropseq.jar PolyATrimmer
OUTPUT_SUMMARY=logs/Surface_naph_polyA_trim.txt MISMATCHES=0 NUM_BASES=5 INPUT=data/Surface_naph_tags_start_filtered_unmapped.bam OUTPUT=data/Surface_naph_trimmed_unmapped.bam
Finished job 3.
4 of 20 steps (20%) done
[Sat Jun 09 16:04:34 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.PolyATrimmer INPUT=data/Surface_naph_tags_start_filtered_unmapped.bam
OUTPUT=data/Surface_naph_trimmed_unmapped.bam OUTPUT_SUMMARY=logs/Surface_naph_polyA_trim.txt MISMATCHES=0 NUM_BASES=5 USE_NEW_TRIMMER=false
TRIM_TAG=ZP ADAPTER=*XX*XCACGTACTCTGCGTTGCTACCACTG MAX_ADAPTER_ERROR_RATE=0.1 MIN_ADAPTER_MATCH=4 MIN_POLY_A_LENGTH=20 MIN_POLY_A_LENGTH_NO_ADAPTER_MATCH=6 DUBIOUS_ADAPTER_MATCH_LENGTH=6 MAX_POLY_A_ERROR_RATE=0.1 VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=0
MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 16:04:34 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
INFO 2018-06-09 16:04:38 TrimStartingSequence Processed 4,000,000 records. Elapsed time: 00:00:58s. Time for last 1,000,000: 1
1s. Last read position: */*
INFO 2018-06-09 16:04:44 PolyATrimmer Processed 1,000,000 records. Elapsed time: 00:00:09s. Time for last 1,000,000: 9s. Last
t read position: */*
INFO 2018-06-09 16:04:52 TrimStartingSequence Processed 5,000,000 records. Elapsed time: 00:01:11s. Time for last 1,000,000: 1
3s. Last read position: */*
INFO 2018-06-09 16:04:53 PolyATrimmer Processed 2,000,000 records. Elapsed time: 00:00:18s. Time for last 1,000,000: 8s. Last
t read position: */*
INFO 2018-06-09 16:05:00 TrimStartingSequence Processed 6,000,000 records. Elapsed time: 00:01:19s. Time for last 1,000,000: 1
8s. Last read position: */*
INFO 2018-06-09 16:05:01 PolyATrimmer Processed 3,000,000 records. Elapsed time: 00:00:26s. Time for last 1,000,000: 8s. Last
t read position: */*
INFO 2018-06-09 16:05:08 TrimStartingSequence Processed 7,000,000 records. Elapsed time: 00:01:28s. Time for last 1,000,000: 1

0s. Last read position: */*
INFO 2018-06-09 16:18:49 TrimStartingSequence Processed 63,000,000 records. Elapsed time: 00:15:08s. Time for last 1,000,000: 1
8s. Last read position: */*
INFO 2018-06-09 16:18:52 PolyATrimmer Processed 59,000,000 records. Elapsed time: 00:14:17s. Time for last 1,000,000: 20s. Last
t read position: */*
INFO 2018-06-09 16:19:00 PolyATrimmer Processed 60,000,000 records. Elapsed time: 00:14:26s. Time for last 1,000,000: 8s. Last
t read position: */*
INFO 2018-06-09 16:19:10 TrimStartingSequence Processed 64,000,000 records. Elapsed time: 00:15:29s. Time for last 1,000,000: 2
0s. Last read position: */*
INFO 2018-06-09 16:19:18 TrimStartingSequence Processed 65,000,000 records. Elapsed time: 00:15:38s. Time for last 1,000,000: 2
8s. Last read position: */*
INFO 2018-06-09 16:19:21 PolyATrimmer Processed 61,000,000 records. Elapsed time: 00:14:47s. Time for last 1,000,000: 20s. Last
t read position: */*
INFO 2018-06-09 16:19:30 PolyATrimmer Processed 62,000,000 records. Elapsed time: 00:14:55s. Time for last 1,000,000: 8s. Last
t read position: */*
INFO 2018-06-09 16:19:39 TrimStartingSequence Processed 66,000,000 records. Elapsed time: 00:15:59s. Time for last 1,000,000: 2
0s. Last read position: */*
INFO 2018-06-09 16:19:48 TrimStartingSequence Processed 67,000,000 records. Elapsed time: 00:16:07s. Time for last 1,000,000: 2
8s. Last read position: */*
INFO 2018-06-09 16:19:50 PolyATrimmer Processed 63,000,000 records. Elapsed time: 00:15:16s. Time for last 1,000,000: 20s. Last
t read position: */*
INFO 2018-06-09 16:20:05 PolyATrimmer Number of reads trimmed: 13908299
INFO 2018-06-09 16:20:05 PolyATrimmer Number of reads completely trimmed: 157072
[Sat Jun 09 16:20:06 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.PolyATrimmer done. Elapsed time: 15.52 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Surface_naph_tags_start_filtered_unmapped.bam.
Finished job 14.
6 of 20 steps (30%) done

localrule plot_polyA_trim:
  input: logs/Surface_naph_polyA_trim.txt
  output: plots/Surface_naph_polyA_trimmed.pdf
  jobid: 2
  wildcards: sample=Surface_naph

rule sam_to_fastq:
  input: data/Surface_naph_trimmed_unmapped.bam
  output: data/Surface_naph_trimmed_unmapped.fastq.gz
  jobid: 26
  wildcards: sample=Surface_naph

[Sat Jun 09 16:20:09 EDT 2018] picard.sam.SamToFastq INPUT=data/Surface_naph_trimmed_unmapped.bam FASTQ=/dev/stdout COMPRESSION_LEVEL=0 OUTPUT=/dev/stdout
T_PER_RG=false RG_TAG=PU REVERSE=true INTERLEAVE=false INCLUDE_NON_PF_READS=false CLIPPING_MIN_LENGTH=0 READ1_TRIM=0 READ2_TRIM=0 INCLUDE_NON_PF_READS=false
PRIMARY_ALIGNMENTS=false VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false
GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 16:20:09 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 2.10.3-SNAPSHOT
Loading required package: magrittr

```

```

rule sam_to_fastq:
    input: data/Surface_naph_trimmed_unmapped.bam
    output: data/Surface_naph_trimmed_unmapped.fastq.gz
    jobid: 26
    wildcards: sample=Surface_naph

[Sat Jun 09 16:20:09 EDT 2018] picard.sam.SamToFastq INPUT=data/Surface_naph_trimmed_unmapped.bam FASTQ=/dev/stdout COMPRESSION_LEVEL=0 OUTPUT=
T PER RG=false RG_TAG=PU RE REVERSE=true INTERLEAVE=false INCLUDE_NON_PF_READS=false CLIPPING_MIN_LENGTH=0 READ1_TRIM=0 READ2_TRIM=0 INCLUDE_NON_PF_READS=
PRIMARY_ALIGNMENTS=false VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 16:20:09 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 2.10.3-SNAPSHOT
Loading required package: magrittr
`geom_smooth()` using method = 'loess'
Finished job 2.
7 of 20 steps (35%) done
INFO 2018-06-09 16:20:12 TrimStartingSequence Processed 68,000,000 records. Elapsed time: 00:16:31s. Time for last 1,000,000: 4s. Last read position: */*
INFO 2018-06-09 16:20:20 TrimStartingSequence Processed 69,000,000 records. Elapsed time: 00:16:40s. Time for last 1,000,000: 8s. Last read position: */*
INFO 2018-06-09 16:20:25 SamToFastq Processed 1,000,000 records. Elapsed time: 00:00:15s. Time for last 1,000,000: 15s. Last read position: */*
INFO 2018-06-09 16:20:28 TrimStartingSequence Number of reads trimmed: 3033867 total reads: 69013983
[Sat Jun 09 16:20:28 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.TrimStartingSequence done. Elapsed time: 16.80 minutes.
Runtime.totalMemory()=1485307904
Removing temporary output file data/Gland_homeo_tags_filtered_unmapped.bam.
Finished job 25.
8 of 20 steps (40%) done

rule polyA_trim:
    input: data/Gland_homeo_tags_start_filtered_unmapped.bam
    output: data/Gland_homeo_trimmed_unmapped.bam, logs/Gland_homeo_polyA_trim.txt
    jobid: 24
    wildcards: sample=Gland_homeo

localrule plot_barcode_start_trim:
    input: logs/Gland_homeo_start_trim.txt
    output: plots/Gland_homeo_start_trim.pdf
    jobid: 9
    wildcards: sample=Gland_homeo

```

```

localrule plot_barcode_start_trim:
    input: logs/Gland_homeo_start_trim.txt
    output: plots/Gland_homeo_start_trim.pdf
    jobid: 9
    wildcards: sample=Gland_homeo

+ java -Xmx16g -Djava.io.tmpdir=/data/clintko/DropSeq Lung/tmp -jar /data/clintko/DropSeq Lung/Drop-seq tools-1.13/jar/dropseq.jar PolyATrimmer
OUTPUT SUMMARY=logs/Gland_homeo_polyA_trim.txt MISMATCHES=0 NUM_BASES=5 INPUT=data/Gland_homeo_tags_start_filtered_unmapped.bam OUTPUT=data/Gland_homeo_trimmed_unmapped.bam
[Sat Jun 09 16:20:29 EDT 2018] org.broadinstitute.dropseqrna.readtrimming.PolyATrimmer INPUT=data/Gland_homeo_tags_start_filtered_unmapped.bam OUTPUT=data/Gland_homeo_trimmed_unmapped.bam OUTPUT SUMMARY=logs/Gland_homeo_polyA_trim.txt MISMATCHES=0 NUM_BASES=5 USE_NEW_TRIMMER=false TRIMMING_METHOD=ZP ADAPTER=XXM*XCACGTACTCTGGTTGCTACCACTG MAX_ADAPTER_ERROR_RATE=0.1 MIN_ADAPTER_MATCH=4 MIN_POLY_A_LENGTH=20 MIN_POLY_A_LENGTH_NO_ADAPTER_MATCH=6 DUBIOUS_ADAPTER_MATCH_LENGTH=6 MAX_POLY_A_ERROR_RATE=0.1 VERBOSITY=INFO QUIET=false VALIDATION_STRINGENCY=STRICT COMPRESSION_LEVEL=5 MAX_RECORDS_IN_RAM=500000 CREATE_INDEX=false CREATE_MD5_FILE=false GA4GH_CLIENT_SECRETS=client_secrets.json
[Sat Jun 09 16:20:29 EDT 2018] Executing as clintko@phl-biostats-ub1 on Linux 4.13.0-43-generic amd64; OpenJDK 64-Bit Server VM 1.8.0_171-8u171-b11-0ubuntu0.17.10.1-b11; Picard version: 1.13(7bed8f4_1513008033)
Loading required package: magrittr
Finished job 9.
9 of 20 steps (45%) done
INFO 2018-06-09 16:20:38 PolyATrimmer Processed 1,000,000 records. Elapsed time: 00:00:09s. Time for last 1,000,000: 9s. Last read position: */*
INFO 2018-06-09 16:20:39 SamToFastq Processed 2,000,000 records. Elapsed time: 00:00:29s. Time for last 1,000,000: 14s. Last read position: */*
INFO 2018-06-09 16:20:47 PolyATrimmer Processed 2,000,000 records. Elapsed time: 00:00:17s. Time for last 1,000,000: 8s. Last read position: */*
INFO 2018-06-09 16:20:53 SamToFastq Processed 3,000,000 records. Elapsed time: 00:00:42s. Time for last 1,000,000: 13s. Last read position: */*
INFO 2018-06-09 16:20:55 PolyATrimmer Processed 3,000,000 records. Elapsed time: 00:00:26s. Time for last 1,000,000: 8s. Last read position: */*
INFO 2018-06-09 16:21:06 SamToFastq Processed 4,000,000 records. Elapsed time: 00:00:56s. Time for last 1,000,000: 13s. Last read position: */*
INFO 2018-06-09 16:21:20 SamToFastq Processed 5,000,000 records. Elapsed time: 00:01:10s. Time for last 1,000,000: 13s. Last read position: */*
INFO 2018-06-09 16:21:28 PolyATrimmer Processed 4,000,000 records. Elapsed time: 00:00:59s. Time for last 1,000,000: 33s. Last read position: */*
INFO 2018-06-09 16:21:33 SamToFastq Processed 6,000,000 records. Elapsed time: 00:01:23s. Time for last 1,000,000: 13s. Last read position: */*
INFO 2018-06-09 16:21:40 PolyATrimmer Processed 5,000,000 records. Elapsed time: 00:01:11s. Time for last 1,000,000: 11s. Last read position: */*
INFO 2018-06-09 16:21:47 SamToFastq Processed 7,000,000 records. Elapsed time: 00:01:36s. Time for last 1,000,000: 13s. Last read position: */*
INFO 2018-06-09 16:21:48 PolyATrimmer Processed 6,000,000 records. Elapsed time: 00:01:19s. Time for last 1,000,000: 8s. Last read position: */*
INFO 2018-06-09 16:21:56 PolyATrimmer Processed 7,000,000 records. Elapsed time: 00:01:27s. Time for last 1,000,000: 8s. Last read position: */*

```



```

t read position: */*
INFO 2018-06-09 16:33:31 PolyATrimmer Processed 56,000,000 records. Elapsed time: 00:13:02s. Time for last 1,000,000: 20s. Las
t read position: */*
INFO 2018-06-09 16:33:38 SamToFastq Processed 59,000,000 records. Elapsed time: 00:13:28s. Time for last 1,000,000: 13s. Las
t read position: */*
INFO 2018-06-09 16:33:40 PolyATrimmer Processed 57,000,000 records. Elapsed time: 00:13:10s. Time for last 1,000,000: 8s. Las
t read position: */*
INFO 2018-06-09 16:33:52 SamToFastq Processed 60,000,000 records. Elapsed time: 00:13:42s. Time for last 1,000,000: 13s. Las
t read position: */*
INFO 2018-06-09 16:34:00 PolyATrimmer Processed 58,000,000 records. Elapsed time: 00:13:31s. Time for last 1,000,000: 20s. Las
t read position: */*
INFO 2018-06-09 16:34:06 SamToFastq Processed 61,000,000 records. Elapsed time: 00:13:56s. Time for last 1,000,000: 13s. Las
t read position: */*
INFO 2018-06-09 16:34:08 PolyATrimmer Processed 59,000,000 records. Elapsed time: 00:13:39s. Time for last 1,000,000: 8s. Las
t read position: */*
INFO 2018-06-09 16:34:16 PolyATrimmer Processed 60,000,000 records. Elapsed time: 00:13:47s. Time for last 1,000,000: 8s. Las
t read position: */*
INFO 2018-06-09 16:34:19 SamToFastq Processed 62,000,000 records. Elapsed time: 00:14:09s. Time for last 1,000,000: 13s. Las
t read position: */*
INFO 2018-06-09 16:34:32 SamToFastq Processed 63,000,000 records. Elapsed time: 00:14:22s. Time for last 1,000,000: 13s. Las
t read position: */*
INFO 2018-06-09 16:34:37 PolyATrimmer Processed 61,000,000 records. Elapsed time: 00:14:07s. Time for last 1,000,000: 20s. Las
t read position: */*
[Sat Jun 09 16:34:40 EDT 2018] picard.sam.SamToFastq done. Elapsed time: 14.52 minutes.
Runtime.totalMemory()-1485307904
INFO 2018-06-09 16:34:45 PolyATrimmer Processed 62,000,000 records. Elapsed time: 00:14:16s. Time for last 1,000,000: 8s. Las
t read position: */*
Finished job 26.
10 of 20 steps (50%) done

rule trim_single:
  input: data/Surface_naph_trimmed_unmapped.fastq.gz
  output: data/Surface_naph_filtered.fastq.gz
  log: logs/Surface_naph_trimlog.txt
  jobid: 10
  wildcards: sample=Surface_naph
  threads: 10

```

Another error occurred

```

rule trim_single:
  input: data/Surface_naph_trimmed_unmapped.fastq.gz
  output: data/Surface_naph_filtered.fastq.gz
  log: logs/Surface_naph_trimlog.txt
  jobid: 10
  wildcards: sample=Surface_naph
  threads: 10

Traceback (most recent call last):
  File "/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/scripts/ksf8usqu.wrapper.py", line 17, in <module>
    shell("trimmomatic SE {snakemake.params.extra} "
  File "/local_data/env-py3/lib/python3.6/site-packages/snakemake/shell.py", line 130, in __new__
    raise sp.CalledProcessError(retcode, cmd)
subprocess.CalledProcessError: Command 'set -euo pipefail; trimmomatic SE -threads 10 data/Surface_naph_trimmed_unmapped.fastq.gz data/Surface_naph_filtered.fastq.gz LEADING:3 TRAILING:3 SLIDINGWINDOW:4:20 MINLEN:20 ILLUMINACLIP:/data/clintko/DropSeq_Lung/Adapter/NexteraPE-PE.fa:2:30:1
0 > logs/Surface_naph_trimlog.txt 2>&1 ' returned non-zero exit status 127.
Error in rule trim_single:
  jobid: 10
  output: data/Surface_naph_filtered.fastq.gz
  log: logs/Surface_naph_trimlog.txt

RuleException:
  CalledProcessError in line 184 of /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk:
  Command 'set -euo pipefail; /local_data/env-py3/bin/python3 /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/scripts/ksf8usqu.wrapper.py ' returned non-zero exit status 1.
  File "/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk", line 184, in __rule_trim_single
  File "/usr/lib/python3.6/concurrent/futures/thread.py", line 56, in run

```

```

Removing temporary output file data/Gland_homeo_tags_start_filtered_unmapped.bam.
Finished job 24.
11 of 20 steps (55%) done
Shutting down, this might take some time.
Exiting because a job execution failed. Look above for error message
Complete log: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/log/2018-06-09T154738.396279.snakemake.log
(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ trimmomatic
trimmomatic: command not found

```

I uploaded the binary of trimmomatic, a flexible read trimming tool for Illumina NGS data.

The official maintainer is Anthony Bolger and the official web page is here:

<http://www.usadellab.org/cms/index.php?page=trimmomatic>

## Downloading Trimmomatic

Version 0.38: [binary](#), [source](#) and [manual](#)

Version 0.36: [binary](#) and [source](#)

however, in the folder it contains the jar file, to run the trimmomatic, the command is

```
shell("trimmomatic SE {snakemake.params.extra} "  
      "{snakemake.input} {snakemake.output} "  
      "{trimmer} " "{log}")
```

To see how it works: <https://snakemake-wrappers.readthedocs.io/en/stable/>

```
rule trim_single:
  input:
    'data/{sample}_trimmed_unmapped.fastq.gz'
  output:
    data='data/{sample}_filtered.fastq.gz'
  log:
    'logs/{sample}_trimlog.txt'
  params:
    trimmer=[ 'LEADING:{ } TRAILING:{ } SLIDINGWINDOW:{ }:{ } MINLEN:{ } ILLUMINACLIP:{ }:{ }:{ }'.Format(
      config['FILTER']['trimomatic'], 'LEADING',
      config['FILTER']['trimomatic'], 'TRAILING',
      config['FILTER']['trimomatic'], 'SLIDINGWINDOW', 'windowSize',
      config['FILTER']['trimomatic'], 'SLIDINGWINDOW', 'requiredQuality',
      config['FILTER']['trimomatic'], 'MINLEN',
      config['FILTER']['trimomatic'], 'adapters-File',
      config['FILTER']['trimomatic'], 'ILLUMINACLIP', 'seedMismatches',
      config['FILTER']['trimomatic'], 'ILLUMINACLIP', 'palindromeClipThreshold',
      config['FILTER']['trimomatic'], 'ILLUMINACLIP', 'simpleClipThreshold' )],
    extra="--threads 10"
  threads: 10
  wrapper:
    '0.21.0/bio/trimomatic/se'
```

```
shell("trimmomatic SE {snakemake.params.extra} "  
      "{snakemake.input} {snakemake.output} "  
      "{trimmer} " "{log}")
```

below is in my notebook

```
In [17]: trimmomatic-0.38.jar

bash: /data/clintko/Trimmomatic/Trimmomatic-0.38/trimmomatic-0.38.jar: Permission denied

solution from Code Repository for BMMB 852\_Fall 2014

In [18]: cd /data/clintko/Trimmomatic/Trimmomatic-0.38

In [19]: ls

adapters  LICENSE  trimmomatic-0.38.jar

In [20]: echo '#!/bin/bash' > trimmomatic
echo 'java -jar /data/clintko/Trimmomatic-0.38/trimmomatic-0.38.jar $@' >> trimmomatic
chmod +x trimmomatic

In [21]: trimmomatic

Usage:
    PE [-version] [-threads <threads>] [-phred33|-phred64] [-trimlog <trimLogFile>] [-summary <statsSummaryFile>] [-quiet]
    [-validatePairs] [-basein <inputBase>] | <inputFile> <inputFile2>] [-baseout <outputBase>] | <outputFile1P> <outputFile1U> <outp
    utFile2P> <outputFile2U>] <trimmer1>...
    or:
    SE [-version] [-threads <threads>] [-phred33|-phred64] [-trimlog <trimLogFile>] [-summary <statsSummaryFile>] [-quiet] <
    inputFile> <outputFile> <trimmer1>...
    or:
    -version
```

It works!!!!

```

clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ trimmomatic
Usage: PE [-version] [-threads <threads>] [-phred33|-phred64] [-trimlog <trimLogFile>] [-summary <statsSummaryFile>] [-quiet] [-validatePairs]
-basein <inputBase> | <inputFile1> <inputFile2>] [-baseout <outputBase> | <outputFile1P> <outputFile1U> <outputFile2P> <outputFile2U>] <trimmer
>...
or:
SE [-version] [-threads <threads>] [-phred33|-phred64] [-trimlog <trimLogFile>] [-summary <statsSummaryFile>] [-quiet] <inputFile> <outp
tFile> <trimmer1>...
or:
-version
clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$

```

Everything works now except for multiqc command, which I could not install now

```

localrule multiqc_trimmomatic:
  input: logs/Surface_naph_trimlog.txt, logs/Gland_homeo_trimlog.txt
  output: reports/filter.html
  jobid: 6

/bin/bash: multiqc: command not found
Traceback (most recent call last):
  File "/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/scripts/xthdtas5.wrapper.py", line 24, in <module>
    "multiqc"
  File "/local_data/env-py3/lib/python3.6/site-packages/snakemake/shell.py", line 130, in __new__
    raise sp.CallProcessError(retcode, cmd)
subprocess.CallProcessError: Command 'set -euo pipefail; multiqc -m trimmomatic --force -o reports -n filter.html logs ' returned non-zero e
xit status 127.
Error in rule multiqc_trimmomatic:
  jobid: 6
  output: reports/filter.html

RuleException:
CalledProcessError in line 269 of /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk:
Command 'set -euo pipefail; /local_data/env-py3/bin/python3 /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/scripts/xthd
tas5.wrapper.py' returned non-zero exit status 1.
  File "/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/rules/filter.smk", line 269, in __rule_multiqc_trimmomatic
  File "/usr/lib/python3.6/concurrent/futures/thread.py", line 56, in run
Finished job 24.
6 of 9 steps (67%) done
Shutting down, this might take some time.
Exiting because a job execution failed. Look above for error message
Complete log: /data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe/.snakemake/log/2018-06-09T174041.880337.snakemake.log
(env-py3) clintko@phl-biostats-ub1:/data/clintko/DropSeq_Lung/Drop-seq_tools-1.13/dropSeqPipe$ which pip

```

since multiqc is just to create reports and it is the last step of the filter.smk, I will keep moving to the next step: mapping

```

rule plot_BC_drop:
  input:
    Cell_tagged=expand('logs/{sample}_CELL_barcode.txt', sample=samples.index),
    UMI_tagged=expand('logs/{sample}_UMI_barcode.txt', sample=samples.index),
    reads_left=expand('logs/{sample}_reads_left.txt', sample=samples.index),
    trimmomatic_filtered=expand('logs/{sample}_reads_left_trin.txt', sample=samples.index)

  params:
    Cell_length=config['FILTER']['cell-barcode']['end'] - config['FILTER']['cell-barcode']['start']+1,
    UMI_length=config['FILTER']['UMI-barcode']['end'] - config['FILTER']['UMI-barcode']['start']+1,
    min_num_below_Cell=config['FILTER']['cell-barcode']['num-below-quality'],
    min_num_below_UMI=config['FILTER']['UMI-barcode']['num-below-quality'],
    min_Cell_quality=config['FILTER']['cell-barcode']['min-quality'],
    min_UMI_quality=config['FILTER']['UMI-barcode']['min-quality'],
    sample_names=lambda wildcards: samples.index,
    batches=lambda wildcards: samples.loc[samples.index, 'batch']

  conda: '../envs/plots.yaml'
  output:
    pdf='plots/BC_drop.pdf'
  script:
    '../scripts/plot_BC_drop.R'

rule multiqc_trimmomatic:
  input:
    expand('logs/{sample}_trinlog.txt', sample=samples.index)
  params:
    '-m trimmomatic'
  output:
    html='reports/filter.html'
  wrapper:
    '0.21.0/bio/multiqc'

~
~
~
~
~

```