

1. Adapter trimming:

```
for i in `cat list.txt`
do
    echo $(date)
    echo ${i}
    fastp -i ${i}_R1_001.fastq.gz -o ${i}_R1out.fastq.gz -I
    ${i}_R3_001.fastq.gz -O ${i}_R2out.fastq.gz --unpaired1
    ${i}_unP1 --unpaired2 ${i}_unP2
    --adapter_sequence=AGATCGGAAGAGCACACGTCTGAACTCCAGTCA
    --adapter_sequence_r2=AGATCGGAAGAGCGTCGTGTAGGGAAAGAGTGT -p -w 4
    -L --overlap_len_require 25 > ${i}_out.log
    mv fastp.html ${i}.html
    mv fastp.json ${i}.json
done
FastP.sh (END)
```

2. ALIGNMENT

```
bwa mem
/u/home/m/mmorsell/project-collaboratory/cfDNA/index/bwa/GRCh38.PRI_1
ambda ${i}_L001_R1out.fastq.gz ${i}_L001_R2out.fastq.gz >
${i}_bwaMem1.sam
```

#SAM to BAM and first filter

```
awk '{if($7 == "=") print}' ${i}_bwaMem1.sam | samtools view -@ 16
-bS - -o ${i}_bwaMem1.bam
```

#post-alignment sorting and filtering (it can be combined with the previous command):

```
samtools view -@ 16 -h -f 2 -F 524 ${i}_bwaMem1.bam | samtools sort
-@ 16 - -o ${i}_sortedFilter.bam
```

#post filtering index and do qualimap

```
samtools index ${i}_sortedFilter.bam

qualimap bamqc -bam ${i}_sortedFilter.bam -ip --java-mem-size=8G -c
-outdir ${i}_bamqc
```

3. SRSLYUMI-bamtag

```
srslyumi-bamtag --binary -o ${i}.rx.bam --take-fragment 0
${i}_sortedFilter.bam
```

4. index the resulting rx.bam files (samtools index)

5. Umi_tools

```
umi_tools group --output-bam --stdin=${i}.rx.bam --stdout=${i}.bx.bam  
--group-out=${i}_withdups.umicorrection.tsv --extract-umi-method=tag  
--umi-tag=RX --method=directional --paired --unmapped-reads=use
```

6. sort and index the .bx.bam files (samtools sort + index)

7. remove duplicates with Picard

```
java -jar /u/home/m/mmorsell/project-mcdb/install/picard.jar  
MarkDuplicates BARCODE_TAG=BX VALIDATION_STRINGENCY=LENIENT  
I=${i}.bx_sorted.bam O=${i}.bx_RmDup.bam M=${i}rmdup.metrics  
REMOVE_DUPLICATES=TRUE
```

#to select by size:

```
alignmentSieve -b XXX_sortedFilter.bam --filterMetrics log.txt -p 8  
--minFragmentLength 25 --maxFragmentLength 99 -o XXX_short.bam  
alignmentSieve -b XXX_sortedFilter.bam --filterMetrics log.txt -p 8  
--minFragmentLength 100 --maxFragmentLength 500 -o XXX_long.bam
```