**Set-up data and software environment**

**1.copy data and script**  
  
**cp -r /share/lect/nanopore/ ./**

**2.prepare singularity environment**

**singularity build --sandbox singularity\_img/nanopore docker://yaozhong/nanopore\_analysis**

**Basecalling using Guppy:**

Singularity nanopore:~> **time guppy\_basecaller -i data/fast5/lambda -s output/basecall/lambda -c /opt/ont/guppy/data/dna\_r9.4.1\_450bps\_fast.cfg --cpu\_threads\_per\_caller 24**

ONT Guppy basecalling software version 3.2.2+9fe0a78

config file: /opt/ont/guppy/data/dna\_r9.4.1\_450bps\_fast.cfg

model file: /opt/ont/guppy/data/template\_r9.4.1\_450bps\_fast.jsn

input path: data/fast5/lambda

save path: output/basecall/lambda

chunk size: 1000

chunks per runner: 160

records per file: 4000

num basecallers: 1

cpu mode: ON

threads per caller: 24

Found 4000 fast5 files to process.

Init time: 1162 ms

0% 10 20 30 40 50 60 70 80 90 100%

|----|----|----|----|----|----|----|----|----|----|

\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*

Caller time: 151790 ms, Samples called: 310564469, samples/s: 2.04601e+06

Finishing up any open output files.

Basecalling completed successfully.

real 2m39.700s

user 33m16.379s

sys 0m4.588s

**Assembly:**

**1.Read-to-read mapping**

Singularity nanopore:~> **minimap2 -x ava-ont -k15 -w5 output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq > output/assembly/lambda.paf**

[M::mm\_idx\_gen::1.379\*0.95] collected minimizers

[M::mm\_idx\_gen::1.788\*1.39] sorted minimizers

[M::main::1.788\*1.39] loaded/built the index for 4000 target sequence(s)

[M::mm\_mapopt\_update::1.866\*1.37] mid\_occ = 498

[M::mm\_idx\_stat] kmer size: 12; skip: 5; is\_hpc: 0; #seq: 4000

[M::mm\_idx\_stat::1.920\*1.36] distinct minimizers: 1562804 (58.78% are singletons); average occurrences: 5.893; average spacing: 2.978

[M::worker\_pipeline::92.227\*2.93] mapped 4000 sequences

[M::main] Version: 2.17-r954-dirty

[M::main] CMD: minimap2 -x ava-ont -k12 -w5 output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq

[M::main] Real time: 92.320 sec; CPU: 269.897 sec; Peak RSS: 0.751 GB

**2.Contig generation:**

Singularity nanopore:~> **miniasm -f output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq output/assembly/lamabda.paf > output/assembly/lambda\_contig.gfa**

[M::main] ===> Step 1: reading read mappings <===

[M::ma\_hit\_read::1.611\*1.00] read 1912033 hits; stored 2539584 hits and 3403 sequences (26512301 bp)

[M::main] ===> Step 2: 1-pass (crude) read selection <===

[M::ma\_hit\_sub::1.957\*1.00] 3368 query sequences remain after sub

[M::ma\_hit\_cut::1.995\*1.00] 2522551 hits remain after cut

[M::ma\_hit\_flt::2.040\*1.00] 2395243 hits remain after filtering; crude coverage after filtering: 416.04

[M::main] ===> Step 3: 2-pass (fine) read selection <===

[M::ma\_hit\_sub::2.200\*1.00] 3366 query sequences remain after sub

[M::ma\_hit\_cut::2.236\*1.00] 2374010 hits remain after cut

[M::ma\_hit\_contained::2.278\*1.00] 54 sequences and 494 hits remain after containment removal

[M::main] ===> Step 4: graph cleaning <===

[M::ma\_sg\_gen] read 162 arcs

[M::main] ===> Step 4.1: transitive reduction <===

[M::asg\_arc\_del\_trans] transitively reduced 39 arcs

[M::asg\_arc\_del\_multi] removed 0 multi-arcs

[M::asg\_arc\_del\_asymm] removed 1 asymmetric arcs

[M::main] ===> Step 4.2: initial tip cutting and bubble popping <===

[M::asg\_cut\_tip] cut 18 tips

[M::asg\_pop\_bubble] popped 0 bubbles and trimmed 0 tips

[M::main] ===> Step 4.3: cutting short overlaps (3 rounds in total) <===

[M::asg\_arc\_del\_multi] removed 0 multi-arcs

[M::asg\_arc\_del\_asymm] removed 4 asymmetric arcs

[M::asg\_arc\_del\_short] removed 8 short overlaps

[M::asg\_cut\_tip] cut 8 tips

[M::asg\_pop\_bubble] popped 0 bubbles and trimmed 0 tips

[M::asg\_arc\_del\_short] removed 0 short overlaps

[M::asg\_arc\_del\_short] removed 0 short overlaps

[M::main] ===> Step 4.4: removing short internal sequences and bi-loops <===

[M::asg\_cut\_internal] cut 0 internal sequences

[M::asg\_cut\_biloop] cut 0 small bi-loops

[M::asg\_cut\_tip] cut 0 tips

[M::asg\_pop\_bubble] popped 0 bubbles and trimmed 0 tips

[M::main] ===> Step 4.5: aggressively cutting short overlaps <===

[M::asg\_arc\_del\_short] removed 0 short overlaps

[M::main] ===> Step 5: generating unitigs <===

[M::main] Version: 0.3-r179

[M::main] CMD: miniasm -f output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq output/assembly/lamabda.paf

[M::main] Real time: 2.337 sec; CPU: 2.339 sec

**3. Read-contig alignment**

Singularity nanopore:~> **minimap2 output/assembly/lambda\_contig.fasta output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq > output/assembly/lambda\_read-contig.paf**

[M::mm\_idx\_gen::0.008\*1.67] collected minimizers

[M::mm\_idx\_gen::0.015\*2.22] sorted minimizers

[M::main::0.015\*2.22] loaded/built the index for 1 target sequence(s)

[M::mm\_mapopt\_update::0.017\*2.12] mid\_occ = 3

[M::mm\_idx\_stat] kmer size: 15; skip: 10; is\_hpc: 0; #seq: 1

[M::mm\_idx\_stat::0.018\*2.05] distinct minimizers: 8941 (99.69% are singletons); average occurrences: 1.003; average spacing: 5.372

[M::worker\_pipeline::0.758\*2.45] mapped 4000 sequences

[M::main] Version: 2.17-r954-dirty

[M::main] CMD: minimap2 output/assembly/lambda\_contig.fasta output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq

[M::main] Real time: 0.764 sec; CPU: 1.862 sec; Peak RSS: 0.041 GB

1. **Polish**

Write the following command in a shell file.

Run in your home fold:

sh script/polish.sh

if [ ! -d “output/assembly/polish” ]; then

mkdir “output/assembly/polish”

# cp the previous contig as the initial start file

**cp** output/assembly/lambda\_contig.fasta output/assembly/polish/lambda\_contig\_polish\_0.fasta

for i in `seq 1 10`

do

echo “Polish iteration $i”

**minimap2** output/assembly/polish/lambda\_contig\_polish\_$((i-1)).fasta output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq > output/assembly/polish/lambda\_read-contig\_polish\_$((i-1)).paf

**racon** output/basecall/lambda/fastq\_runid\_12b8e1a5cfeaeb8afe4cc42710993b762f9fdae5\_0\_0.fastq output/assembly/polish/lambda\_read-contig\_polish\_$((i-1)).paf output/assembly/polish/lambda\_contig\_polish\_$((i-1)).fasta > output/assembly/polish/lambda\_contig\_polish\_$((i)).fasta

done

**Structural variant detection**

**Base-calling**

**guppy\_basecaller -i data/fast5/hs\_chr20\_1000/ -s output/basecall/hs\_chr20\_1000 -c /opt/ont/guppy/data/dna\_r9.4.1\_450bps\_fast.cfg --cpu\_threads\_per\_caller 24**

**Merge fastq file**

cat output/basecall/hs\_chr20\_1000/\*.fastq > output/basecall/hs\_chr20\_1000/hs\_chr20\_1000.fastq

**NGMLR**

Singularity nanopore:~> **ngmlr -t 24 -x ont -r data/reference/chr20.fa -q output/basecall/hs\_chr20\_1000/hs\_chr21\_1000.fastq -o output/sv/hs\_chr21\_1000.bam**

ngmlr 0.2.8 (build: Oct 4 2019 11:44:29, start: 2019-10-08.05:44:06)

Contact: fritz.sedlazeck@gmail.com, philipp.rescheneder@gmail.com

Opening for output (SAM): output/sv/hs\_chr21\_1000.bam

Encoding reference sequence.

Size of reference genome 64 Mbp (max. 68719 Mbp)

0 reference sequences were skipped (length < 10).

Writing encoded reference to data/reference/chr20.fa-enc.2.ngm

Writing to disk took 0.03s

Building reference index #0 (kmer length: 13, reference skip: 2)

854 prefixes were ignored due to the frequency cutoff (1000)

Overall time for creating RefTable: 4.45s

Writing reference index to data/reference/chr20.fa-ht-13-2.2.ngm

Writing to disk took 0.61s

Opening query file output/basecall/hs\_chr20\_1000/hs\_chr21\_1000.fastq

Mapping reads...

Processed: 893 (0.60), R/S: 49.61, RL: 10469, Time: 2.94 2.50 10.89, Align: 0.99, 410, 0.97

Done (536 reads mapped (60.02%), 357 reads not mapped, 984 lines written)(elapsed: 0m, 22 r/s)

**Sort sam file and transform bam**

samtools sort -T output/tmp -o output/sv/hs\_chr21\_1000.sorted.bam output/sv/hs\_chr21\_1000.sam

**Sniffer**

Singularity nanopore:~> **sniffles -m output/sv/hs\_chr21\_1000.sorted.bam -v output/sv/hs\_chr21\_1000.vcf**

Estimating parameter...

Max dist between aln events: 5

Max diff in window: 50

Min score ratio: 2

Avg DEL ratio: 0.0636924

Avg INS ratio: 0.0408543

Start parsing... chr20

Finalizing ..

Start genotype calling:

Reopening Bam file for parsing coverage

Finalizing ..