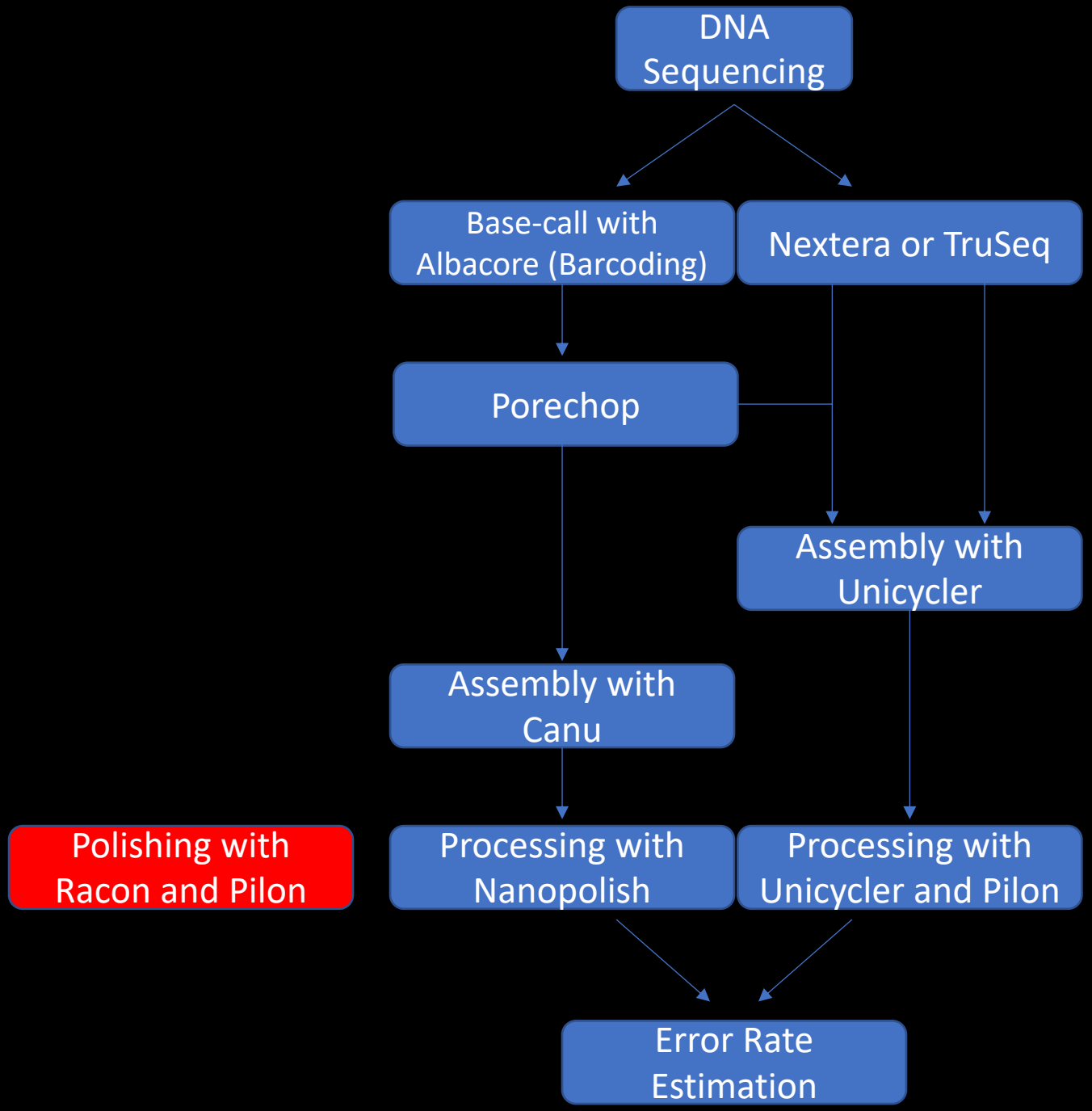
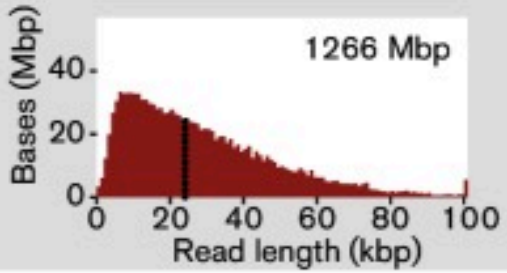
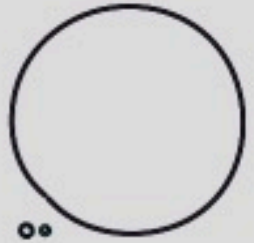
Completing bacterial genome assemblies with multiplex MinION sequencing

Ryan R Wick et al. 2017

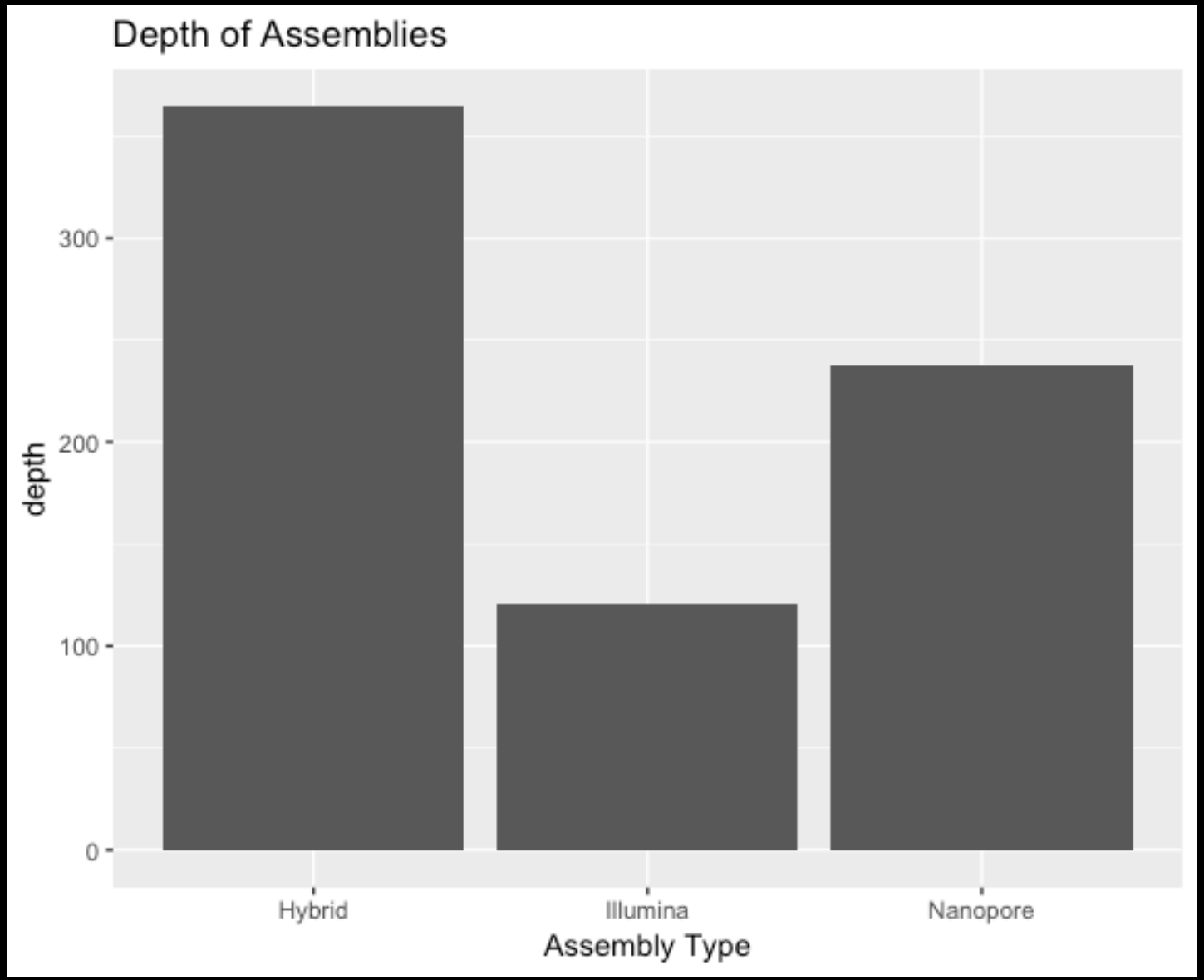

DNA Sequencing
Base-call with Albacore (Barcoding)
Nextera or TruSeq
Porechop
Assembly with Unicycler
Assembly with Canu
Polishing with Racon and Pilon
Processing with Nanopolish
Processing with Unicycler and Pilon
Error Rate Estimation

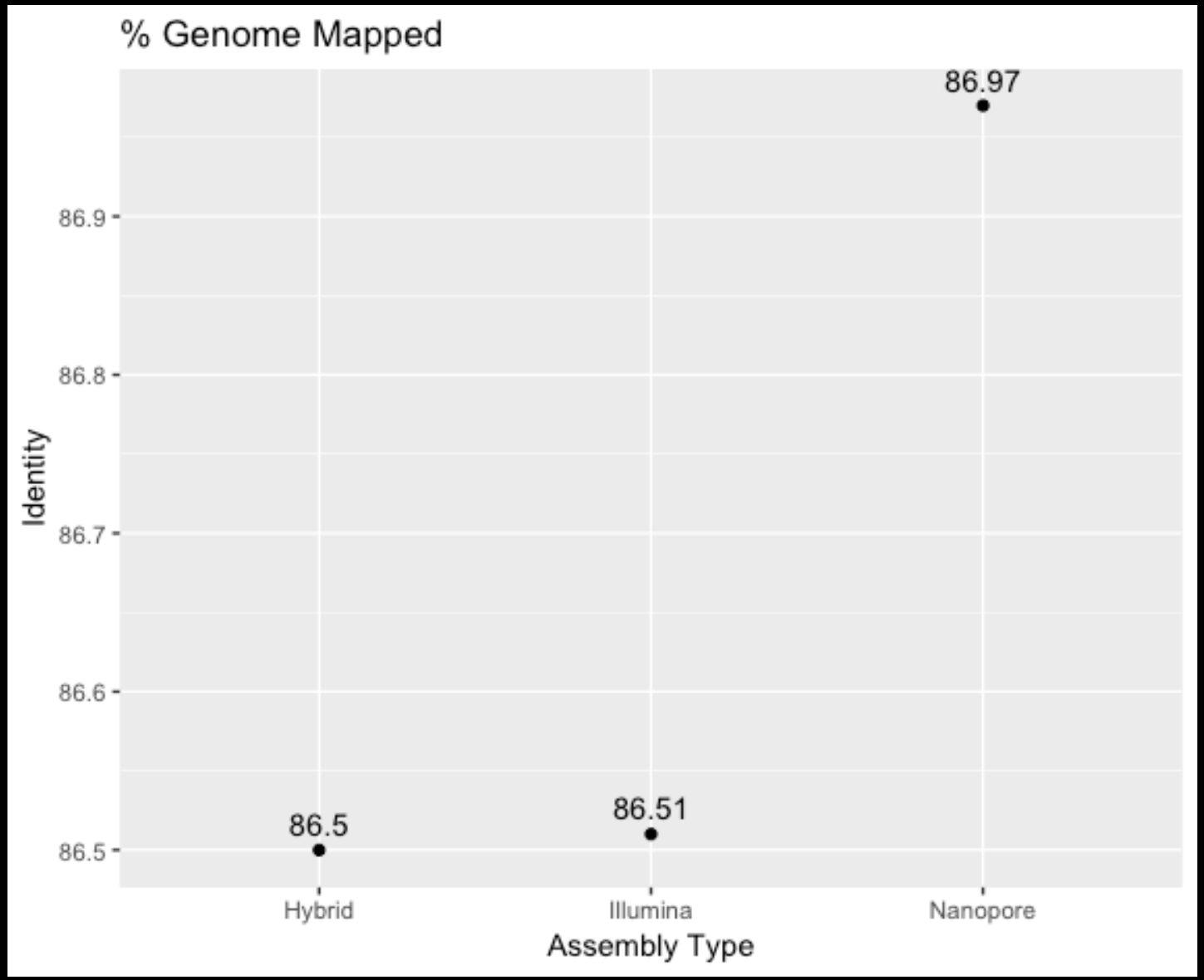
Sample	Illumina-only assembly graph	ONT reads	Hybrid assembly graph
INF042			

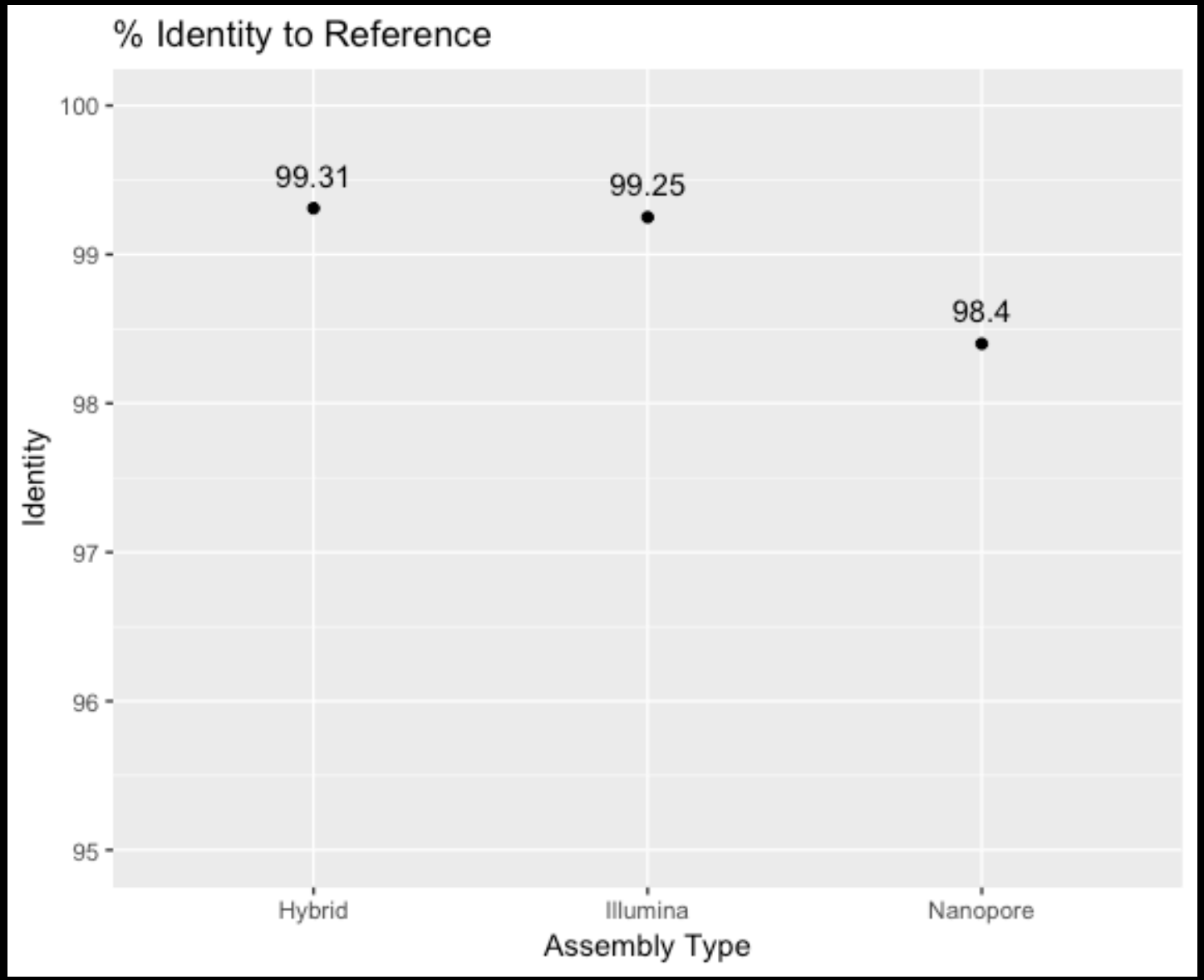
```
refdata=/home/abaryiames/Lab_Projects/polishtest/GCF_000240185.1_ASM24018v2_genomic.fna
reads=/usr/share/data/proj_data/pneumonia/5170843/barcode01.fastq
i_1=/usr/share/data/proj_data/pneumonia/5170831/barcode01_1.fastq.gz
i_2=/usr/share/data/proj_data/pneumonia/5170831/barcode01_2.fastq.gz
/all_segments.fasta
sample_name="barcode01"
mkdir $sample_name
cd $sample_name

logfile='log.txt'
touch $logfile
echo $datapath > $logfile

ls -lh $datapathO

#Porechop - Gets rid of adapter seqs
porechop -i $datapathO/$sample_name* -o barcode01_chop.fastq

chopped_pore=/home/abaryiames/Lab_Projects/pneumonia_script/barcode01_chop.fastq

#Canu - Assembly
canu -p canu -d canu_dir genomeSize=5.5m maxThreads=8 -nanopore-raw $chopped_pore
oxford_canu=/home/abaryiames/Lab_Projects/pneumonia_script/barcode01/canu_dir/canu.contigs.fasta


#Minimap / Racon - Polishing
minimap -Sw5 -L100 -m0 -t2 $assembly $reads > mapIT01.paf
racon -t 16 $reads mapIT01.paf $oxford_canu > mapIT01.fa
minimap -Sw5 -L100 -m0 -t2 mapIT01.fa $reads > mapIT02.paf
racon -t 16 $reads mapIT02.paf mapIT01.fa > mapIT02.fa
minimap -Sw5 -L100 -m0 -t2 mapIT02.fa $reads > mapIT03.paf
racon -t 16 $reads mapIT03.paf mapIT02.fa > mapIT03.fa


#Hybrid Assembly
unicycler -1 $i_1 -2 $i_2 -l $chopped_pore -o unicycler_hybrid_assemblies --threads 16

#Illumina Assembly
unicycler -1 $i_1 -2 $i_2 -o unicycler_illumina_assemblies --threads 16
```

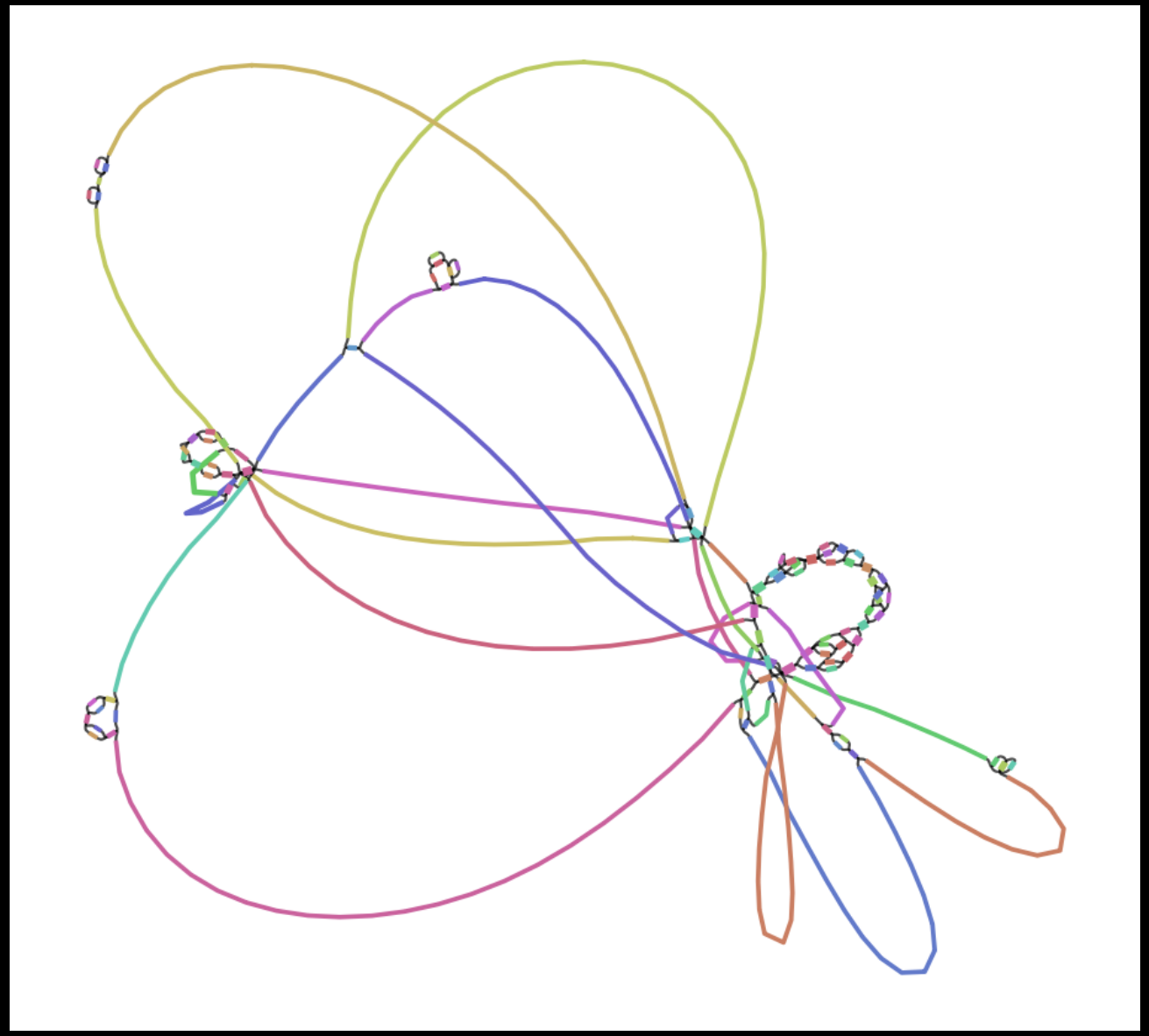
```
#Name assemblies
assemblyO=/home/abaryiames/Lab_Projects/polishtest/mapIT03.fa
assemblyI=/home/abaryiames/Lab_Projects/pneumonia_script/unicycler_illumina_assemblies/assembly.f
asta

assemblyH=/home/abaryiames/Lab_Projects/pneumonia_script/unicycler_hybrid_assemblies/read_align
ment/all_segments.fasta

#dnadiff
dnadiff -p oxford_pilon $refdata $pilon
dnadiff -p oxford_pilon $refdata $assemblyI
dnadiff -p oxford_pilon $refdata $assemblyH

#Quast.py
quast.py $assemblyO -R $refdata --nanopore $reads -o quast_oxford --threads 16
quast.py $assemblyI -R $refdata --pe1 $i_1 --pe2 $i_2 -o quast_illumina --threads 16
quast.py $assemblyH -R $refdata --pe1 $i_1 --pe2 $i_2 --nanopore $reads -o quast_hybrid --threads 16
```

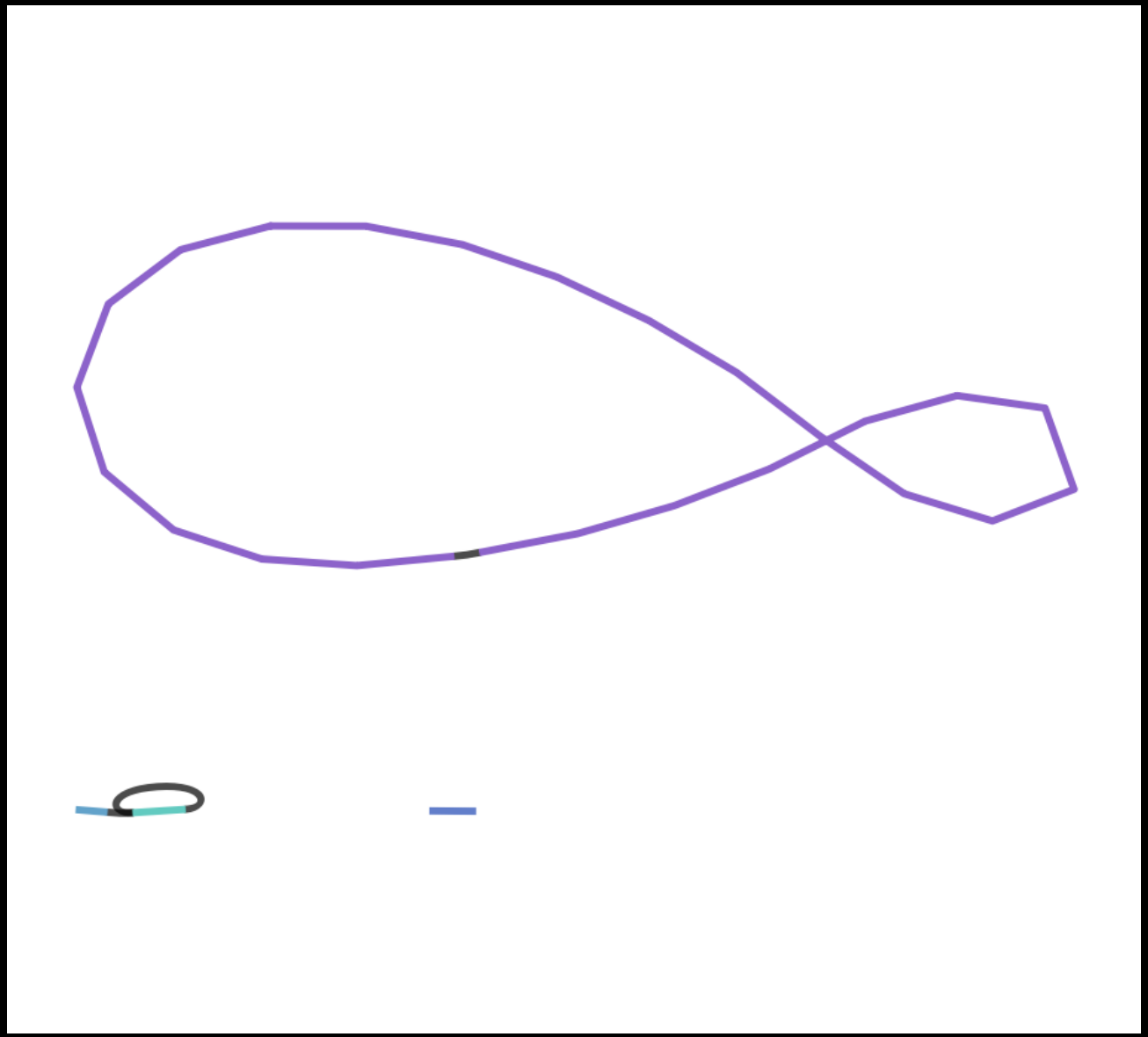

Read lengths vs Average read quality plot
Average read quality
Read lengths
0
2
4
6
8
10
12
14
0
20000
40000
60000
80000
100000
120000
140000

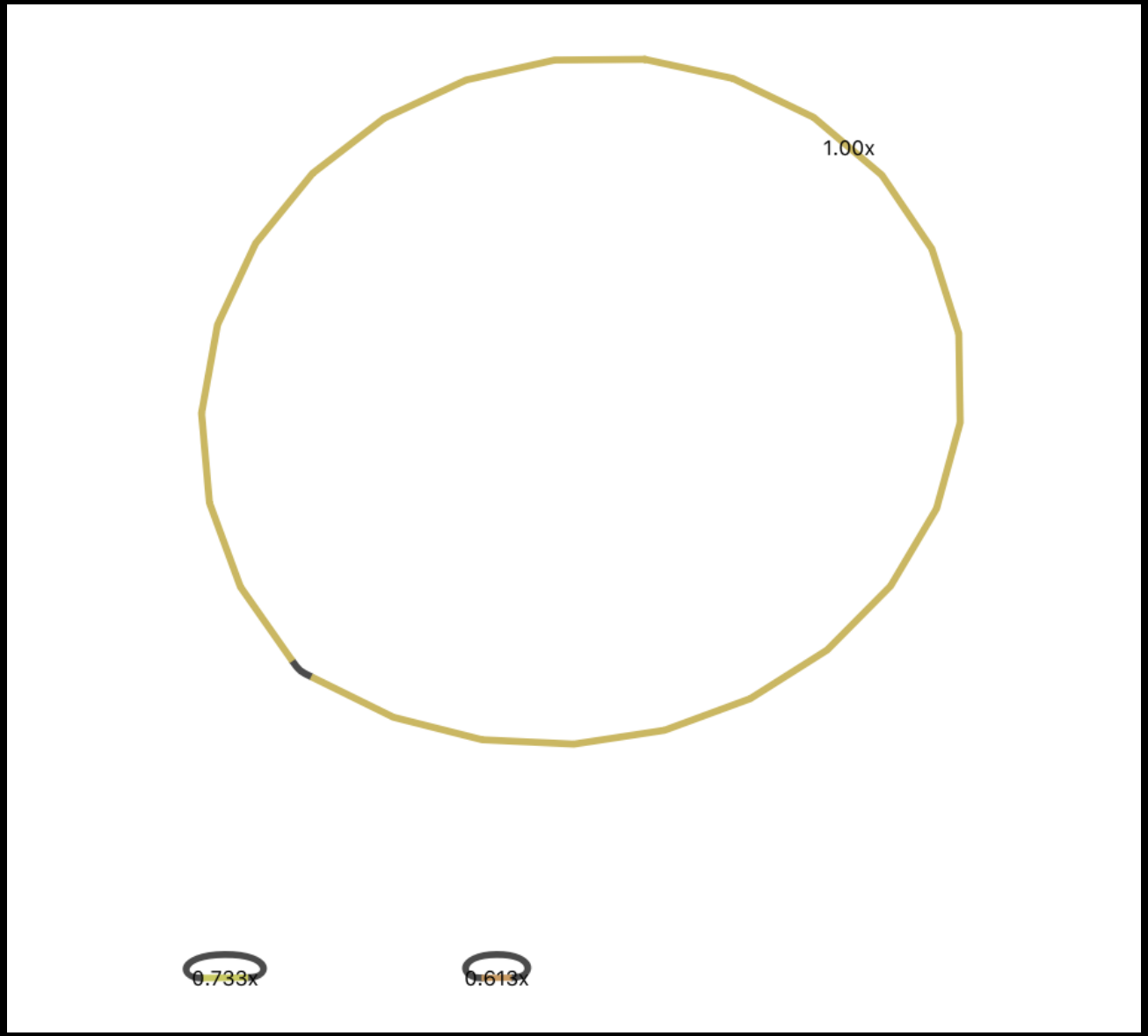

Depth of Assemblies
depth
300
200
100
0
Hybrid
Illumina
Nanopore
Assembly Type


% Genome Mapped
86.97
86.9
86.8
86.7
86.6
86.5
Identity
86.5
86.51
Hybrid
Illumina
Nanopore
Assembly Type


% Identity to Reference
100
99
98
97
96
95
Identity
99.31
99.25
98.4
Hybrid
Illumina
Nanopore
Assembly Type






1.00x
0.733x
0.613x