Afbeelding met tekst

Automatisch gegenereerde beschrijving

**Your partner for**

**Innovative OMICS**

**solutions**

**DRY-LAB NOTES**

Project number : OHMX20230016R

Customer :

Date :

Author : Steven Verbruggen, Arne Godderis

**Table of contents**

[1 Sample info 3](#_Toc103698507)

[2 Test run 3](#_Toc103698508)

[3 Full run 3](#_Toc103698509)

[4 References 3](#_Toc103698510)

# Sample info

4 samples were multiplexed for adaptive sampling on PromethION.

|  |  |  |
| --- | --- | --- |
| Sample ID on tube | OHMX ID | Barcode |
| SY-SY5Y 2/9/22 R1 | OHMX20230016R\_001 | 1 |
| IMR-32 14/11/22 R1 | OHMX20230016R\_002 | 2 |
| SKNDE2C 14/11/22 R1 | OHMX20230016R\_003 | 3 |
| SHEP 12/12/22 R1 | OHMX20230016R\_004 | 4 |

During the first reload the flowcell got loose from the device causing the sequencing run to not be stopped properly, this is why there is no minknow html report present for the first “run”.

# Readfish + QC

Toml file was configured for readfish, file can be found at:

"G:\Shared drives\OHMX.bio\PROJECTS\DRYLAB\internalResearch\OHMX20230016R\_Ferroptosis\promethion\_FRGs.toml"

Readfish targets --toml promethion\_FRGs.toml --device 2F --log OHMX20230016R\_rf.log --experiment-name OHMX20230016R

Performed by Arne

# Sup basecalling + mods

Basecalling:

* SUP model for pore R10.4.1
* Calling of 5mC and 5hmC in CG context
* Barcode detection based on barcoding kit NBD114 (24 barcodes possible, 4 used (1-4) in wet lab)
* Min score (parameter --min\_score) was left at default 0

#For first loading

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0)

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0)\_5mCG\_5hmCG@v3.1

dorado basecaller -r ./dna\_r10.4.1\_e8.2\_400bps\_sup\@v4.2.0 ../pod5/ --modified-bases 5mCG\_5hmCG --kit-name SQK-NBD114-24 > OHMX20230016R\_Ferroptosis\_no\_sample.calls.bam

#For reloaded flow cell data

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0)

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0) \_5mCG\_5hmCG@v3.1

dorado basecaller -r ./dna\_r10.4.1\_e8.2\_400bps\_sup\@v4.2.0 ../pod5/ --modified-bases 5mCG\_5hmCG --kit-name SQK-NBD114-24 > OHMX20230016R\_Ferroptosis\_reload.calls.bam

kit-name will add codes for detected barcodes. These can later be divided in files per barcode using ‘dorado demux’

Divide into files per barcode. Classification of barcodes is done earlier.

dorado demux --output-dir demux\_output/ --no-classify OHMX20230016R\_Ferroptosis\_no\_sample.calls.bam

dorado demux --output-dir demux\_output/ --no-classify OHMX20230016R\_Ferroptosis\_reload.calls.bam

To keep the modified basecalling info: map with the dorado aligner (based on minimap2) instead of minimap2

Concatenate first load and reloaded data (samtools merge for sorted BAM files)

Mapping, BAM file sorting, merging and index is all done in loop script for all barcodes:

declare -A datasets\_first\_load=(["barcode01"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/no\_sample/20231114\_1524\_2F\_PAQ18472\_825fbfcd/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode01.bam"

["barcode02"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/no\_sample/20231114\_1524\_2F\_PAQ18472\_825fbfcd/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode02.bam"

["barcode03"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/no\_sample/20231114\_1524\_2F\_PAQ18472\_825fbfcd/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode03.bam"

["barcode04"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/no\_sample/20231114\_1524\_2F\_PAQ18472\_825fbfcd/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode04.bam")

declare -A datasets\_reload=(["barcode01"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/reload/20231115\_1525\_2F\_PAQ18472\_bd7b6b8d/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode01.bam"

["barcode02"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/reload/20231115\_1525\_2F\_PAQ18472\_bd7b6b8d/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode02.bam"

["barcode03"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/reload/20231115\_1525\_2F\_PAQ18472\_bd7b6b8d/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode03.bam"

["barcode04"]="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/reload/20231115\_1525\_2F\_PAQ18472\_bd7b6b8d/sup\_basecalling/demux\_output/SQK-NBD114-24\_barcode04.bam")

ALIGNDIRFIRSTLOAD="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/no\_sample/20231114\_1524\_2F\_PAQ18472\_825fbfcd/sup\_basecalling/aligned\_output/"

ALIGNDIRRELOAD="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/reload/20231115\_1525\_2F\_PAQ18472\_bd7b6b8d/sup\_basecalling/aligned\_output/"

GENOME="/data/igenomes/Homo\_sapiens/Ensembl/GRCh38/Sequence/Minimap2Index/genome\_readfish.mmi"

CORES=12

##Start loop

for i in "${!datasets\_first\_load[@]}"

do

:

ID=$i

echo -e "START NEW LOOP:\n\t${ID}\n"

cd $ALIGNDIRFIRSTLOAD

FIRSTLOADINPUT=${datasets\_first\_load[$i]}

FIRSTLOADALNBAM="$ALIGNDIRFIRSTLOAD/${ID}\_first\_load.bam"

FIRSTLOADSORTBAM="$ALIGNDIRFIRSTLOAD/${ID}\_first\_load.sorted.bam"

echo -e "Mapping first load\n"

dorado aligner -t $CORES $GENOME $FIRSTLOADINPUT > $FIRSTLOADALNBAM

echo -e "Sorting first load\n"

samtools sort -@ $CORES -o $FIRSTLOADSORTBAM $FIRSTLOADALNBAM

cd $ALIGNDIRRELOAD

RELOADINPUT=${datasets\_reload[$i]}

RELOADALNBAM="$ALIGNDIRRELOAD/${ID}\_reoad.bam"

RELOADSORTBAM="$ALIGNDIRRELOAD/${ID}\_reload.sorted.bam"

echo -e "Mapping reload\n"

dorado aligner -t $CORES $GENOME $RELOADINPUT > $RELOADALNBAM

echo -e "Sorting reload\n"

samtools sort -@ $CORES -o $RELOADSORTBAM $RELOADALNBAM

MERGEDIR="/storage/OHMX20230016R\_Ferroptosis/OHMX20230016R/sup\_concat/"

cd $MERGEDIR

MERGEDOUTPUT="$MERGEDIR/${ID}.sorted.bam"

echo -e "Merge first load and reload\n"

samtools merge -@ $CORES $MERGEDOUTPUT $FIRSTLOADSORTBAM $RELOADSORTBAM

echo -e "Index merged sorted bam\n"

samtools index -@ $CORES $MERGEDOUTPUT

done

##End loop

Data is stored on GridION storage.

Copy for downstream analysis to Midas: /data/lvisser/data\_OHMX20230016R\_20231114

# OHMX20230016R\_MM

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0)

dorado download --model [dna\_r10.4.1\_e8.2\_400bps\_sup@v4.2.0](mailto:dna_r10.4.1_e8.2_400bps_sup@v4.2.0)\_5mCG\_5hmCG@v3.1

dorado basecaller -r ./dna\_r10.4.1\_e8.2\_400bps\_sup\@v4.2.0 ./pod5/ --modified-bases 5mCG\_5hmCG --kit-name SQK-NBD114-24 > ./sup\_basecalling/OHMX20230016R\_MM.calls.bam