ONTrack2 tutorial

This tutorial will describe the steps required for running ONTrack2 pipeline for obtaining consensus sequences.

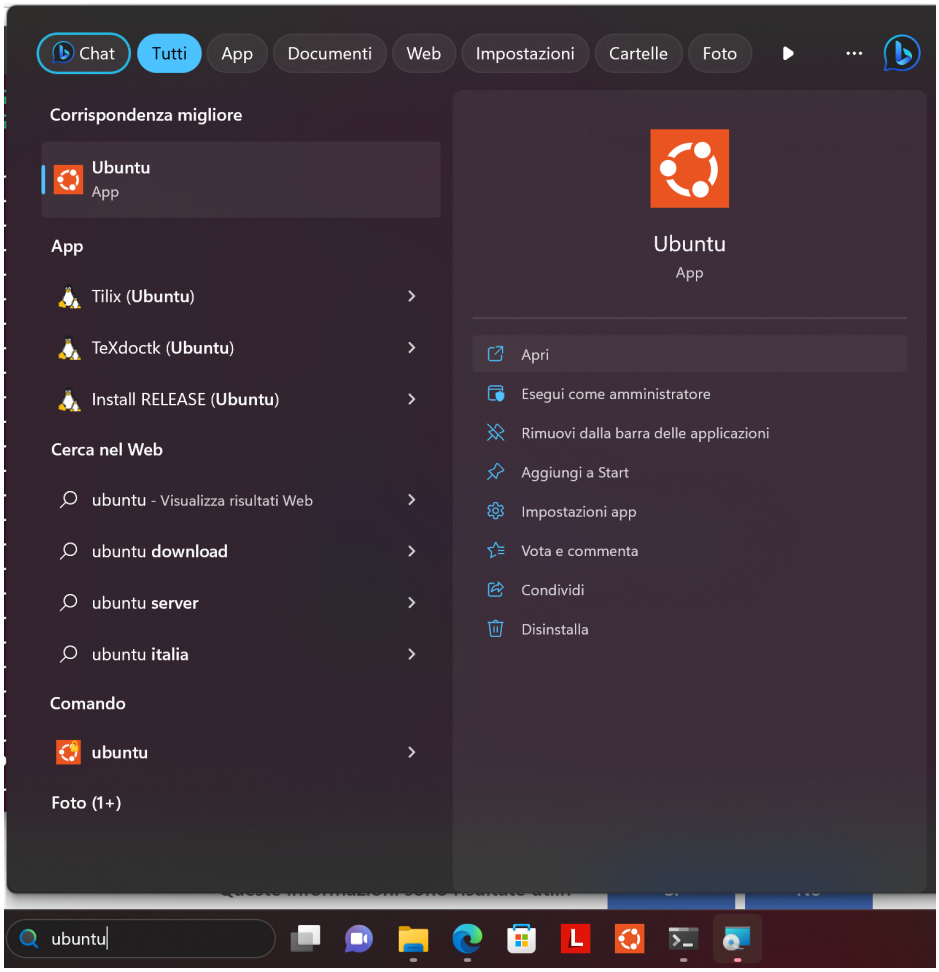
Sequencing of Nanopore reads is performed using MinKNOW software in Windows OS. Right after the sequencing with MinKNOW is done, raw reads are saved in fast5 (or pod5) format. These reads need to be base-called, demultiplexed and concatenated, in order to obtain one fastq file for each sample.

Base-calling and demultiplexing can be performed either "live" by MinKNOW (recommended option) or off-line, for example with ONT_preprocessing pipeline. In the following, the two options will be described.

Option #1: live base-calling and demultiplexing

If you run live base-calling and demultiplexing, you will end up with one folder including a set of subfolders named <barcodexx>, each of them containing a set of fastq files, which need to be concatenated before running **ONTrack2** pipeline, in order to obtain one fastq for each barcode.

1. Start Ubuntu app from Windows, searching for Ubuntu in the search bar.



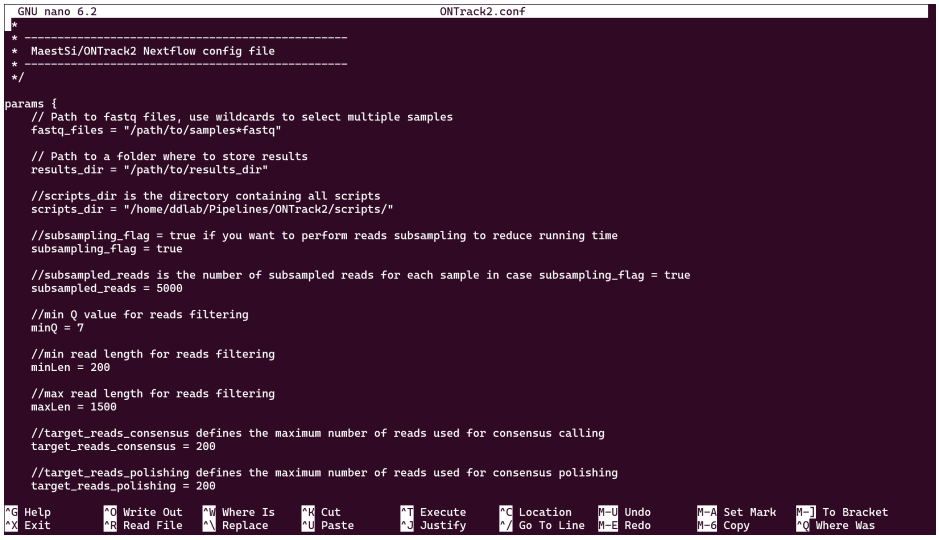
2. Using the Terminal, go to **ONTrack2** pipeline folder, with:

```
cd /home/ddlab/Pipelines/ONTrack2
```

3. Open *ONTrack2.conf* file and set desired parameters with:

```
nano /home/ddlab/Pipelines/ONTrack2/ONTrack2.conf
```

To save/discard changes and exit, press 'Ctrl+x' followed by 'y' or 'n' and 'enter'.

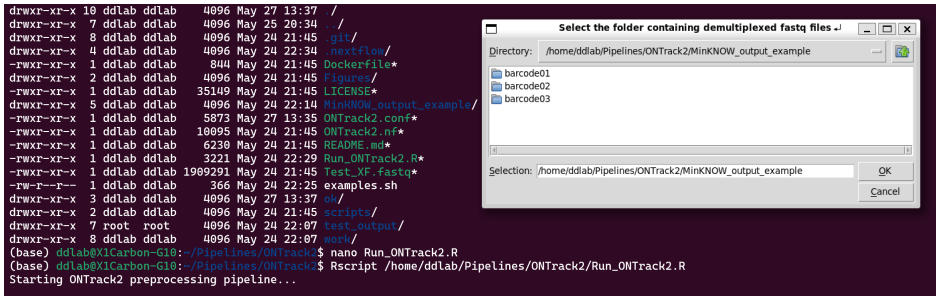


4. Perform concatenation of fastq files and run **ONTrack2** pipeline with:

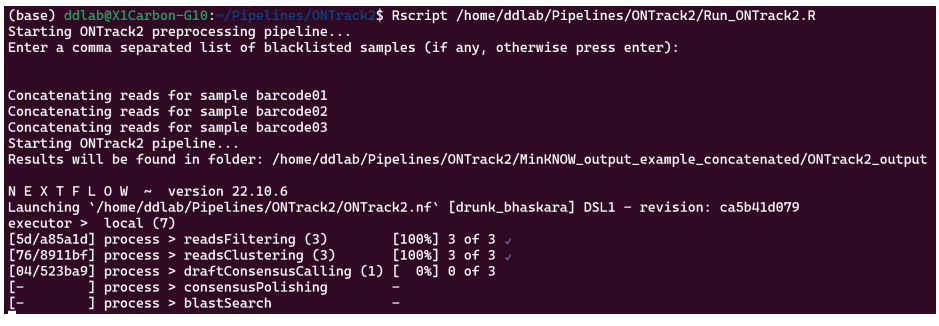
```
Rscript /home/ddlab/Pipelines/ONTrack2/Run_ONTrack2.R
```

C and D hard-drives can be accessed from Ubuntu at the path `/mnt/c` or `/mnt/d` , respectively.

5. Follow the interactive procedure. You will be required to select the folder with fastq files to be concatenated (produced by MinKNOW), a list of barcodes you want to exclude, the *ONTrack2.conf* file (look for it in /home/ddlab/Pipelines/ONTrack2.conf) and the *ONTrack2.nf* file (in the same folder). Be sure to double click on the file/folder you want to select.



Type the list of barcodes you want to exclude (if any), and wait for the results.



6. Look at the folder printed on screen. You will find consensus sequences and Blast search results in blastSearch folder.

```
(base) ddlab@X1Carbon-G10:~/Pipelines/ONTrack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch$ ll /home/ddlab/Pipelines/ONT
rack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch/
total 16
drwxr-xr-x 2 root  root  4096 May 27 13:49  /
drwxr-xr-x 8 ddlab ddlab 4096 May 27 13:49 ../
-rw-r--r-- 1 root  root     0 May 27 13:49 barcode01_blast_results.txt
-rw-r--r-- 1 root  root   931 May 27 13:49 barcode01_consensus.fasta
-rw-r--r-- 1 root  root     0 May 27 13:49 barcode02_blast_results.txt
-rw-r--r-- 1 root  root   931 May 27 13:49 barcode02_consensus.fasta
(base) ddlab@X1Carbon-G10:~/Pipelines/ONTrack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch$
```

You can copy the folder to Windows with:

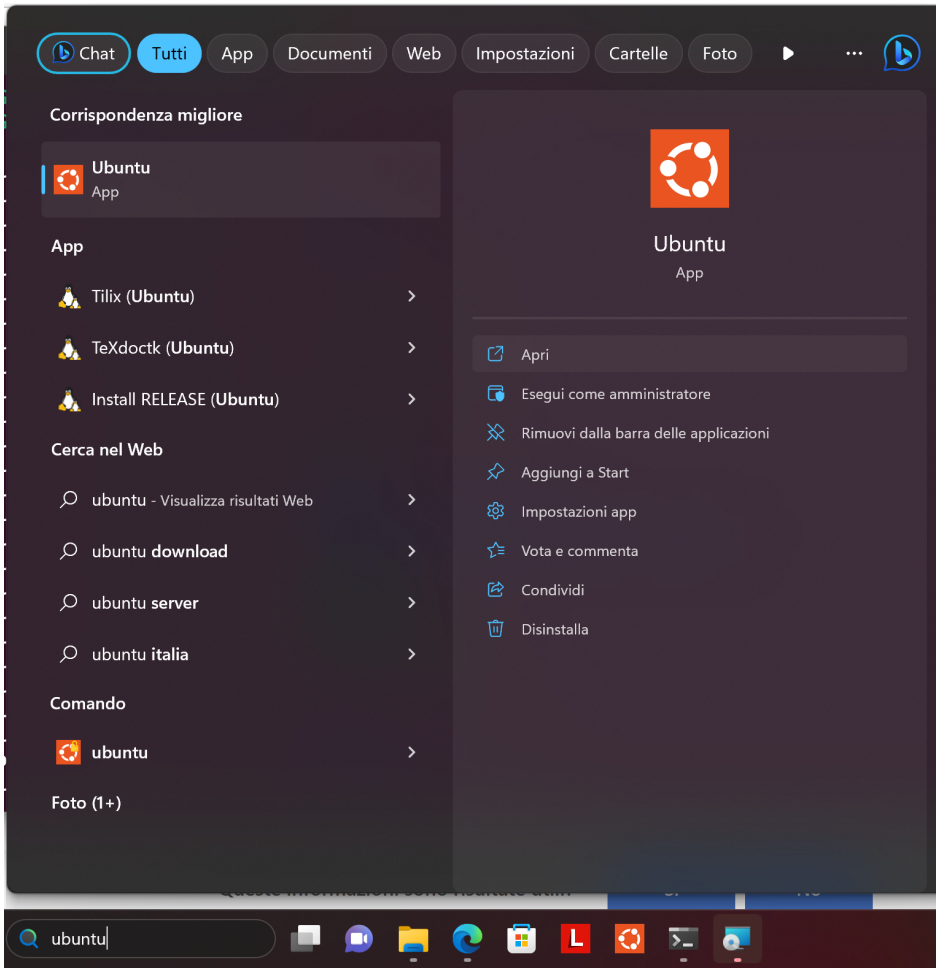
```
cp -r "/full/path/to/results/dir" /mnt/c/Users/DDlab/Desktop/
```

and open the files with a standard text editor.

For more in-depth information, have a look at https://github.com/MaestSi/ONTrack2 repository.

Option #2: offline base-calling and demultiplexing

1. Start Ubuntu app from Windows, searching for Ubuntu in the search bar.



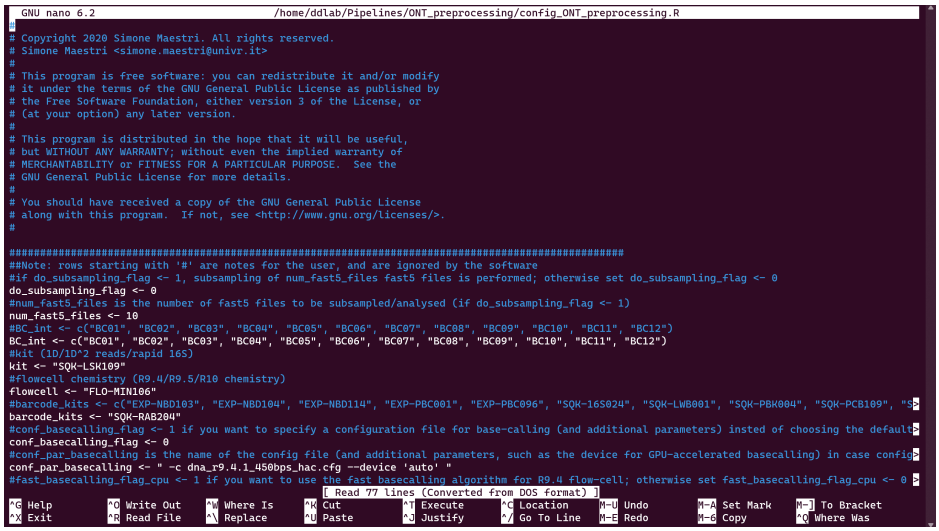
2. Using the Terminal, go to **ONT_preprocessing** pipeline folder:

```
cd /home/ddlab/Pipelines/ONT_preprocessing
```

3. Open *config_ONT_preprocessing.R* and set desired parameters with:

```
nano /home/ddlab/Pipelines/ONT_preprocessing/config_ONT_preprocessing.R
```

To save/discard changes and exit, press 'Ctrl+x' followed by 'y' or 'n' and 'enter'.



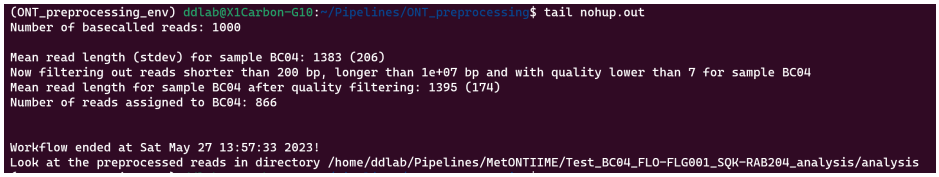
4. Run **ONT_preprocessing** pipeline with:

```
/home/ddlab/Pipelines/ONT_preprocessing/Launch_ONT_preprocessing.sh <fast5 dir>
```

with <fast5 dir> being the path to the folder containing raw fast5 reads. C and D hard-drives can be accessed from Ubuntu at the path `/mnt/c` or `/mnt/d`, respectively.

Wait for base-calling and demultiplexing to complete. Demultiplexed and concatenated fastq files will be found in <fast5 dir>_analysis/analysis folder. You can look at the progress of the process typing:

```
tail /home/ddlab/Pipelines/ONT_preprocessing/nohup.out
```



5. Using the Terminal, go to **ONTrack2** pipeline folder, with:

```
cd /home/ddlab/Pipelines/ONTrack2
```

6. Open *ONTrack2.conf* file and set desired parameters with:

```
nano /home/ddlab/Pipelines/ONTrack2/ONTrack2.conf
```

To save/discard changes and exit, press 'Ctrl+x' followed by 'y' or 'n' and 'enter'.

```
 GNU nano 6.2                                                ONTrack2.conf
/*
* -------------------------------------------------
*  MaestSi/ONTrack2 Nextflow config file
* -------------------------------------------------
*/

params {
    // Path to fastq files, use wildcards to select multiple samples
    fastq_files = "/path/to/samples*fastq"

    // Path to a folder where to store results
    results_dir = "/path/to/results_dir"

    //scripts_dir is the directory containing all scripts
    scripts_dir = "/home/ddlab/Pipelines/ONTrack2/scripts/"

    //subsampling_flag = true if you want to perform reads subsampling to reduce running time
    subsampling_flag = true

    //subsampled_reads is the number of subsampled reads for each sample in case subsampling_flag = true
    subsampled_reads = 5000

    //min Q value for reads filtering
    minQ = 7

    //min read length for reads filtering
    minLen = 200

    //max read length for reads filtering
    maxLen = 1500

    //target_reads_consensus defines the maximum number of reads used for consensus calling
    target_reads_consensus = 200

    //target_reads_polishing defines the maximum number of reads used for consensus polishing
    target_reads_polishing = 200

^G Help        ^O Write Out   ^W Where Is    ^K Cut         ^T Execute     ^C Location    M-U Undo       M-A Set Mark   M-] To Bracket
^X Exit        ^R Read File   ^\ Replace     ^U Paste       ^J Justify     ^/ Go To Line  M-E Redo       M-6 Copy       ^Q Where Was
```

7. Run **ONTrack2** pipeline with:

```
cd /home/ddlab/Pipelines/ONTrack2/
nextflow -c ONTrack2.conf run ONTrack2.nf --
fastq_files="/full/path/to/fastq/folder/*.fastq" --
results_dir="/full/path/to/results/dir" -profile docker
```

8. Look at the results folder. You will find consensus sequences and blast search results in blastSearch folder.

```
(base) ddlab@X1Carbon-G10:~/Pipelines/ONTrack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch$ ll /home/ddlab/Pipelines/ONT
rack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch/
total 16
drwxr-xr-x 2 root  root  4096 May 27 13:49  /
drwxr-xr-x 8 ddlab ddlab 4096 May 27 13:49  /
-rw-r--r-- 1 root  root     0 May 27 13:49 barcode01_blast_results.txt
-rw-r--r-- 1 root  root   931 May 27 13:49 barcode01_consensus.fasta
-rw-r--r-- 1 root  root     0 May 27 13:49 barcode02_blast_results.txt
-rw-r--r-- 1 root  root   931 May 27 13:49 barcode02_consensus.fasta
(base) ddlab@X1Carbon-G10:~/Pipelines/ONTrack2/MinKNOW_output_example_concatenated/ONTrack2_output/blastSearch$
```

You can copy the folder to Windows with:

```
cp -r "/full/path/to/results/dir" /mnt/c/Users/DDlab/Desktop/
```

and open the files with a standard text editor.

For more in-depth information, have a look at https://github.com/MaestSi/ONT_preprocessing and https://github.com/MaestSi/ONTrack2 repositories.