ChIPseq pipeline (Quick start)

(2024-12-16)

#### 1. Setup new directory for analysis

mkdir PROJECTID && cd PROJECTID

where PROJECTID a reasonable name like "G233\_G228\_ATAC\_DAR"

#### 2. Clone directory

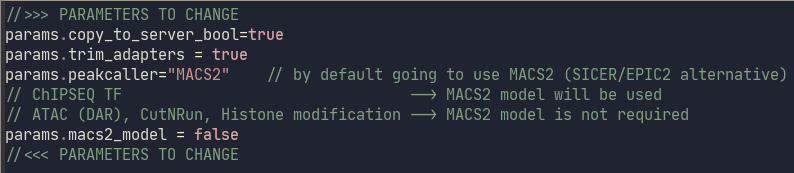
git clone https://github.com/mpyatkov/chipseq\_nextflow.git ./

#### 3. Setup config files

**3.1** copy XLSX configuration file and edit it

cp ./docs/G233\_G228\_ATAC\_DAR\_example.xls ./

**3.2** In addition, you need to check and edit main.nf to configure the following parameters:



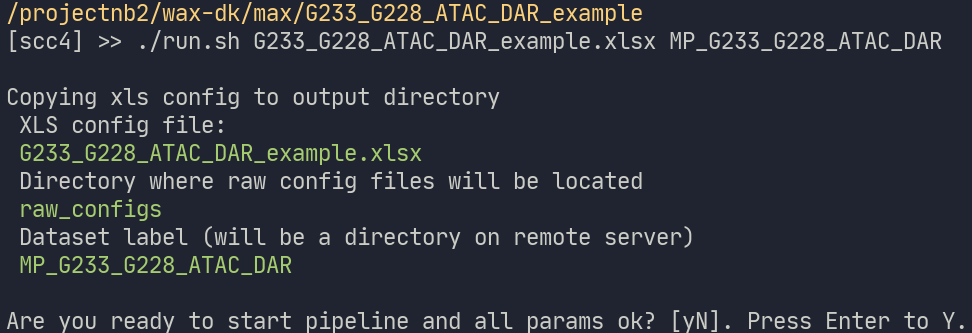
|  |  |  |
| --- | --- | --- |
| **Parameter** | **Value** | **Description** |
| params.copy\_to\_server | **true**/false | manage if we need to copy MULTIQC, BigWig files and tracks to server |
| params.trim\_adapters | **true**/false | Trimming adapters |
| params.peakcaller | **MACS2**/SICER | Select peakcaller for analysis (MACS2 – narrow peaks, SICER – broad peaks) |
| params.macs2\_model | true/**false** | Using model for MACS2. True only for dataset when we do ChIP-seq for transcription factors. False for ATAC (DAR), CutNRun and Histone modification ChIP-seq datasets |

#### 4. Starting the pipeline

To execute the pipeline, use the script below with a parameter corresponding to the dataset label. The dataset label will be included in the results directory name and will also serve as the name of the directory on the Waxman server, which contains the UCSC, BigWig, and BED track files, along with the MULTIQC report.

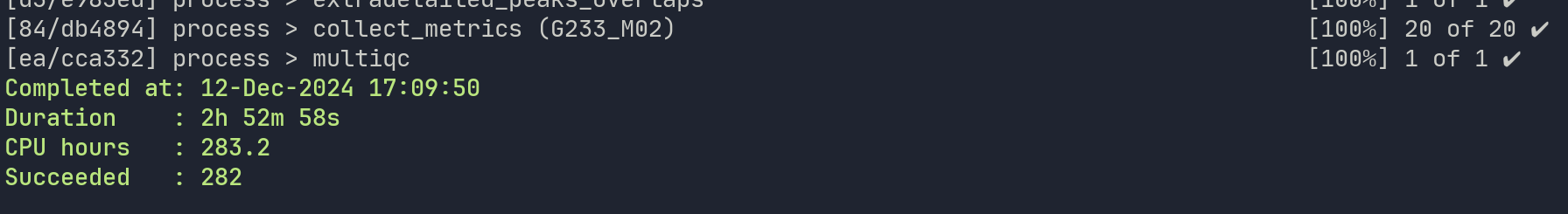
./run.sh ./G233\_G228\_ATAC\_DAR\_example.xls MP\_G233\_G228\_ATAC\_DAR

The script will ask you at the beginning if all the information provided below is correct, if everything is ok, just press Enter to continue.



After pressing Enter, the script will first try to install the R packages, unpack the configurations into the raw\_configs directory (they don't need to be edited), and run nextflow to process the main.nf file.

After finishing all the jobs, you want to see the green message with execution time:



That green light means that everything is ok, and it is time for you to aggregate the information and create a UCSC Browser session.

Setup UCSC sessions

All tracks for user project will be located here:

<https://waxmanlabvm.bu.edu/><**your BU username**>/<**Dataset label**>

Specifically for this example my path will be the following:

[https://waxmanlabvm.bu.edu/**mpyatkov**/**MP\_G233\_G228\_ATAC\_DAR**/](https://waxmanlabvm.bu.edu/mpyatkov/MP_G233_G228_ATAC_DAR/)TRACK\_LINES

We are interested only in these two tracks:

autolimit\_tracks.txt

autolimit\_combined\_tracks.txt

Click on the file to open, then Save As to download a copy of each .txt file. Save in the ChIP-seq pipeline Summary folder, if you wish to retain a local copy.

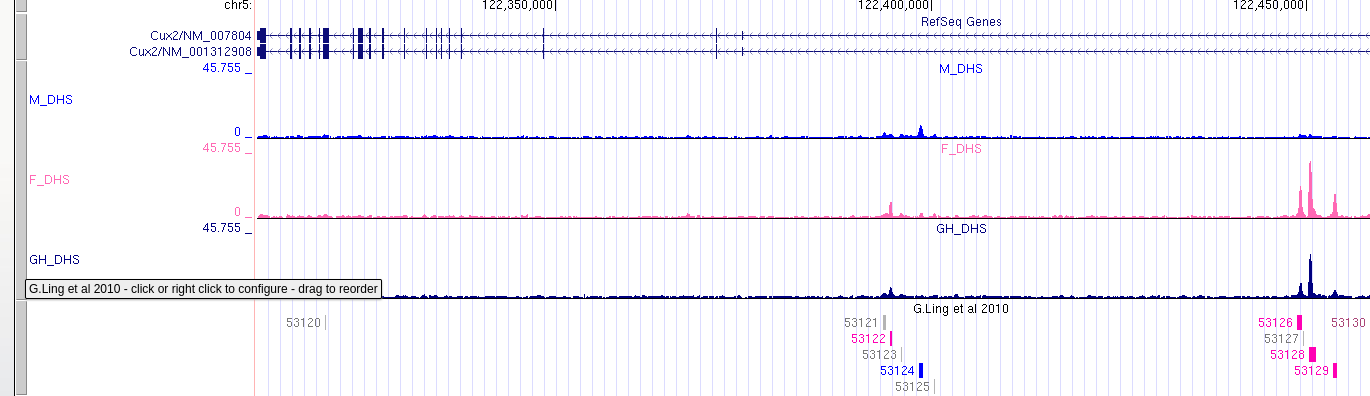
Session preparation

1. Go to the UCSC Browser, and sign in to the lab’s account with prior saved sessions (User Name: Djw). Use DEFAULT\_CHIPSEQ\_SESSION as a base session (you can find it in the list of sessions) using the Search box).

2. We need to create 3 sessions:

1. MP\_G233\_G228\_ATAC\_DAR\_Individual

Open the base session DEFAULT\_CHIPSEQ\_SESSION, and click on **Manage Custom Tracks**. Click on ‘**Add Custom Tracks**’. Click on ‘upload: Browse…’ above the white box, then navigate to the autolitmit\_tracks.txt . Click **Submit**, then click ‘**Go to First Annotation**’ to view the session. Manually drag the following three Reference tracks to the top:

Finally, save the session.

2. MP\_G233\_G228\_ATAC\_DAR\_Individual\_GroupScaled

Use the previously created MP\_G233\_G228\_ATAC\_DAR\_Individual as a base session, go to **My data => Track collection builder.** Click ‘**Add Collection**’ (top right) to make a new folder (and assign a folder name) on the right side. Next, on the left side, manually click on each bright green arrow to add each of the individual tracks to the folder you created on the right side. When all of the tracks that you wish to merge have been moved from the left to the right side, click ‘**Go**’ at the top right to merge all BigWig tracks into one track. This new track should appear as the last track in the current session. Go to **My data => Custom tracks** and delete all BigWig and their corresponding SICER and MACS2 tracks. Configure the merged track by setting the following parameters:

track height: 64

Dataview scaling: group-auto-scale

Always include zero: ON

Draw y indicator lines: at y = 0.0: ON

Finally, save the session, including ‘\_Individual\_GroupScaled’ in the file name.

3. MP\_G233\_G228\_ATAC\_DAR\_Combined\_GroupScaled

Use the DEFAULT\_CHIPSEQ\_SESSION as a base session, then upload and aggregate autolimit\_combined\_tracks.txt as you did for MP\_G233\_G228\_ATAC\_DAR\_Individual\_GroupScaled

Finally, save the session, including ‘\_Combined\_GroupScaled’ in the file name.

Run the UCSC downloader script

Once all UCSC sessions have been created, you can download all top 25 up/downregulated regions for each comparison by running the following commands:

cd download\_ucsc\_pdf

./run.sh “MP\_G233\_G228\_ATAC\_DAR\_Individual\_GroupScaled”

After that, you will have to wait until you see the following text:

All downloads are complete. Check ../RESULTS\_MP\_G233\_G228\_ATAC\_DAR/summary/UCSC\_TOP25\_PDF

You can download pdfs from multiple sessions at once by adding the other sessions as another argument to run.sh

./run.sh “session1” “session2” ...

Summarize information

Ultimately, once you have set up your sessions and uploaded all the data, send DW an email that contains the following information:

1. The link to the OneDrive directory containing all content from the directory RESULTS\_MP\_G233\_G228\_ATAC\_DAR/summary/

2. The links to the three UCSC sessions

3. MultiQC report which will be located here: https://waxmanlabvm.bu.edu/mpyatkov/MP\_G233\_G228\_ATAC\_DAR/multiqc/multiqc\_report.html