BenchMapL

Release 1

Baudeau

Jun 12, 2022

CONTENTS:

CHAPTER

ONE

REQUIREMENT

For the good functionnement of the Workflow please make sure the different package are installed

1.1 Python Module

Import:

- Conda
- Snakemake
- matplotlib
- pysam
- numpy
- upsetplot

1.2 OS version

- Linux

CHAPTER

TWO

API

2.1 Workflow

BenchPlot.**common_error_gp1**(*results*, *output*)

plot upsetplot of unmapped reads groups

Parameters

- **results** (*list*) – list of each resu of the groups
- **output** (*str*) – name of the output

BenchPlot.**common_error_gp2**(*results*, *output*)

plot upsetplot of poorly located reads groups

Parameters

- **results** (*list*) – list of each resu of the groups
- **output** (*str*) – name of the output

BenchPlot.**countdiff**(*bamFP*)

fill the different result for the input bam

Parameters
bamFP (*pybam object*) – bam file in pybam format

Returns
the resu object

Return type
resu

BenchPlot.**find_output**(*key*, *name=False*, *outpath=None*)

find the output for each groups

Parameters

- **key** (*str*) – name of the group
- **name** (*bool or str, optional*) – name of the selected output, defaults to False
- **outpath** (*list, optional*) – list of all the possible output name, defaults to None

Returns
True if outpout is correct else false

Return type
boolean

BenchPlot.**get_MapOrNot**(*results*)

return number of mapped and unmapped reads for each result of a groups

Parameters
results (*list*) – list of resu object

Returns
lists of the number of mapped and unmapped reads for each resu

Return type
list

BenchPlot.**get_all_cor**(*results*)

return list of list of each cor for each result of a groups

Parameters
results (*list*) – list of resu object

Returns
list of list of each cor

Return type
list

BenchPlot.**get_label**(*results*)

extracts name of all the resu object of the groups

Parameters
results (*list*) – list of resu object

Returns
list of name

Return type
list

BenchPlot.**get_nbHuman**(*results*)

return number of missaligned reads for each result of a groups

Parameters
results (*list*) – list of resu object

Returns
list of the number of missaligned reads for each resu

Return type
list

BenchPlot.**is_human**(*read*, *resu*)

tell if read is an human reads and increm missaligned

Parameters

- **read** (*pybam object*) – pybam object
- **resu** (*resu object*) – resu object

Returns
False if missaligned else True

Return type
boolean

BenchPlot.**multiple_cor**(*results*, *output*)

plot of the number of reads correctly localised by categories

Parameters

- **results** (*list*) – list of each resu of the groups
- **output** (*str*) – name of the output

BenchPlot.**parsenamesimu**(*text*)

return expected start position end position and number of reads mapped

Parameters
text (*str*) – name of the read

Returns
expected start position, expected end postion and number of mapped reads

Return type
tupple

BenchPlot.**plot_histoMU**(*results*, *output*)

plot histogram of unmapped /mapped reads

Parameters

- **results** (*list*) – list of each resu of the groups
- **output** (*str*) – name of the output

BenchPlot.**plot_histoNotHuman**(*results*, *output*)

plot histogram of the number of human reads mapped

Parameters

- **results** (*list*) – list of each resu of the groups
- **output** (*str*) – name of the output

BenchPlot.**read_align**(*resu*, *read*, *threshold=0*)

increm resu cor for a read

Parameters

- **resu** (*resu*) – class resu
- **read** (*pybam object*) – read object from pybam
- **threshold** (*int, optional*) – specified cor number, defaults to 0

Returns
result of the operation true if read is increm else false

Return type
boolean

class BenchPlot.**result**

class for keept the different result from a bam

addGroup1(*add*)

add reads in group1 reads (unmapped reads)

Parameters
add (*str*) – id of the unmapped reads

addGroup2(*add*)

add reads in group2 reads (misslocalised reads)

Parameters
add (*str*) – id of the reads

corpercent(*cor*)

percent of cor reads

Parameters
cor (*int*) – number of cor reads

Returns
percent of correct reads

Return type
int

increm_cor(*num=0*)

increm cor

Parameters
num (*int, optional*) – num of the cor, defaults to 0

increm_mapped()

increment mapped

increm_unmapped()

increm mapped

mappercentM()

return percent of mapped reads

Returns
percent of mapped reads

Return type
int

mappercentU()

return percent of unmapped reads

Returns
percent of unmapped reads

Return type
int

missalign()

increm missaligned

setname(*name*)

set name

Parameters
name (*str*) – name of the result

plot_default.**do_something**(*data_path*, *out_path*, *myparam*)

main function

Parameters

- **data_path** (*list*) – snakemake.input
- **out_path** (*list*) – snakemake.output
- **myparam** (*list*) – config files

plot_default.**group_input**(*files*)

generate a dict for each group of tools/datasets

Parameters
files (*list[string]*) – list of bam file from the different tools/dataset

Returns
dictionnary of the different datasets groups with each files in a list

Return type
dict

plot_default.**parsepath**(*path*)

find the name of the tool

Parameters
path (*string*) – a path

Returns
the string between the / and _ character

Return type
string

plot_params.**do_something**(*data_path*, *out_path*, *myparam*)

main function

Parameters

- **data_path** (*list*) – snakemake.input
- **out_path** (*list*) – snakemake.output
- **myparam** (*list*) – config files

plot_params.**group_input**(*files*)

generate a dict for each group of tools/datasets/command

Parameters
files (*list[string]*) – list of bam file from the different tools/dataset

Returns
dictionnary of the different datasets groups with each files in a list

Return type
dict

plot_params.**parsepath**(*path*)

check the command and the tools

Parameters
path (*string*) – a path

Returns
the string name of the command

Return type
string

2.2 Workflow2

select_human_nanosim.**do_something**(*data_path*, *out_path*)

main function for snakemale

Parameters

- **data_path** (*string*) – snakemake.input
- **out_path** (*string*) – snakemake.output

select_human_nanosim.**select_reads**()

select reads and add them in a list

Returns
list of reads

:rtype:list

select_human_pbsim.**do_something**(*data_path*, *out_path*)

main function for snakemale

Parameters

- **data_path** (*string*) – snakemake.input
- **out_path** (*string*) – snakemake.output

select_human_pbsim.**select_reads**()

select reads and add them in a list

Returns
list of reads

:rtype:list

convert_name.**do_something**()

main function for rename all the reads in a fasta file

convert_name.**findr**(*x*)

filter for find the strand of a reads

Parameters
x (*string*) – name of the reads

Returns
F foward, R reverse

Return type
string

convert_name.**openfile**(*file*)

open a file

Parameters
file (*string*) – name of a file

Returns
list of all the lines of the files

Return type
list

PYTHON MODULE INDEX

INDEX