

HiC-Pro does not perform deduplication by default and the option to do is not documented on the github page

<https://github.com/nservant/HiC-Pro>

How to use it ?

First have a look at the help message !

```
HiC-Pro --help
usage : HiC-Pro -i INPUT -o OUTPUT -c CONFIG [-s ANALYSIS_STEP] [-p] [-h] [-v]
Use option -h|--help for more information

HiC-Pro 2.10.0
-----
OPTIONS

-i|--input INPUT : input data folder; Must contains a folder per sample with input
-o|--output OUTPUT : output folder
-c|--conf CONFIG : configuration file for Hi-C processing
[-p|--parallel] : if specified run HiC-Pro on a cluster
[-s|--step ANALYSIS_STEP] : run only a subset of the HiC-Pro workflow; if not speci
mapping: perform reads alignment
proc_hic: perform Hi-C filtering
quality_checks: run Hi-C quality control plots
build_contact_maps: build raw inter/intrachromosomal contact maps
ice_norm: run ICE normalization on contact maps

[-h|--help]: help
[-v|--version]: version
```

config-hicpro.GB.txt

```
#####
## Digestion Hi-C
#####

GENOME_FRAGMENT =
LIGATION_SITE =
MIN_FRAG_SIZE = 100
MAX_FRAG_SIZE = 100000
MIN_INSERT_SIZE = 100
MAX_INSERT_SIZE = 600

#####
## Hi-C processing
#####

MIN_CIS_DIST = 1000
GET_ALL_INTERACTION_CLASSES = 1
GET_PROCESS_SAM = 1
RM_SINGLETON = 1
RM_MULTII = 1
RM_DUP = 1

#####
## Contact Maps
#####

BIN_SIZE = 1000000 500000 200000 100000 40000
MATRIX_FORMAT = complete # upper

#####
## ICE Normalization
#####

MAX_ITER = 100
FILTER_LOW_COUNT_PERC = 0.02
FILTER_HIGH_COUNT_PERC = 0
EPS = 0.1
```

hic-pro/2.7.6/Linux/RHEL6/x86_64/bin/HiC-Pro

```
## This software is distributed without any guarantee under the terms of the GNU General
## Public License, either Version 2, June 1991 or Version 3, June 2007.

SOFT="HiC-Pro"
VERSION="2.7.6"

function usage {
    echo -e "usage : $$SOFT -i INPUT -o OUTPUT -c CONFIG [-s ANALYSIS_STEP] [-p] [-h] [-v]"
    echo -e "Use option -h|--help for more information"
}

function help {
    usage;
    echo
    echo "$$SOFT $$VERSION"
    echo "-----"
    echo "OPTIONS"
    echo
    echo " -i|--input INPUT : input data folder; Must contains a folder per sample with input files"
    echo " -o|--output OUTPUT : output folder"
    echo " -c|--conf CONFIG : configuration file for Hi-C processing"
    echo " [-p|--parallel] : if specified run $$SOFT on a cluster"
    echo " [-s|--step ANALYSIS_STEP] : run only a subset of the $$SOFT workflow; if not specified the"
    echo " mapping: perform reads alignment"
    echo " proc_hic: perform Hi-C filtering"
    echo " quality_checks: run Hi-C quality control plots"
    echo " merge_persample: merge multiple inputs and remove duplicates if specified"
    echo " build_contact_maps: run merge_persample and build raw inter/intrachromosomal contact"
    echo " ice_norm: run ICE normalization on contact maps"
    echo " [-h|--help]: help"
    echo " [-v|--version]: version"
    exit;
}
```

Copy and edit the configuration file 'config-hicpro.txt' in your local folder. See the [manual](#) for details about the configuration file