Hi-C / HiChIP data analysis

Documented by Victoria Noe-Kim

Last updated on June 1, 2018

Table of Contents

[1. HiC-Pro Installation 1](#_Toc515635737)

[1.1 Online manual for HiC-Pro 1](#_Toc515635738)

[1.2 Requirements 1](#_Toc515635739)

[1.2.1 Comments on different versions of HiC-Pro 1](#_Toc515635740)

[1.2.2 Installation of the required dependencies 2](#_Toc515635741)

[a. miniconda2 2](#_Toc515635742)

[b. python libraries: pysam, numpy, scipy, bx 2](#_Toc515635743)

[c. samtools 2](#_Toc515635744)

[d. bowtie 2 3](#_Toc515635745)

[e. unix sort 3](#_Toc515635746)

[1.2.3 How to update or install necessary dependencies 3](#_Toc515635747)

[numpy: 3](#_Toc515635748)

[scipy 3](#_Toc515635749)

[bx (bx-python) 3](#_Toc515635750)

[Installing GNU: 3](#_Toc515635751)

[Homebrew 4](#_Toc515635752)

[2. Running HiC-Pro 4](#_Toc515635753)

[2.1 Generate higher resolution matrix 5](#_Toc515635754)

[3. Post-processing HiC-Pro output files 5](#_Toc515635755)

[3.1 Hicpro2juicebox 5](#_Toc515635756)

[3.1.1 Virtual 4C plot 6](#_Toc515635757)

[3.1.2 Juicebox dump function 7](#_Toc515635758)

[3.2 hicpro2higlass 8](#_Toc515635759)

[3.3 hicpro2fithic 8](#_Toc515635760)

[3.3.1 Fit-Hi-C: Call significant loops 9](#_Toc515635761)

[3.4 Hipro\_macs2 11](#_Toc515635762)

[3.5 Virtual 4C plot for 3D genome browser 12](#_Toc515635763)

[3.6 Hichipper: specific to analyzing HiChIP data 13](#_Toc515635764)

[4. Troubleshoot 13](#_Toc515635765)

[Comments 13](#_Toc515635766)

[4.1 #1 Issue: HiC-Pro runs into an error while running —make:\*\*\*[bowtie\_pairing] Error 1 13](#_Toc515635767)

[4.2 #2 Issue: sort: stray character in field spec: invalid field specification ‘2,2V’ 14](#_Toc515635768)

[4.3 #3 Issue: Interference with iTerm color scheme 14](#_Toc515635769)

# HiC-Pro Installation

## 1.1 Online manual for HiC-Pro

References:

Online manual: <http://nservant.github.io/HiC-Pro/>

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

HiC-Pro forum: <https://groups.google.com/forum/#!forum/hic-pro>

# Requirements

Note

These requirements are also listed in the references listed in section 1.1. These pipelines are constantly updated so it is important to make sure that pipelines and their dependencies are up to date.

1. Bowtie2
2. Python (>2.7, python 3 not support)
   1. pysam (>= 0.8.3)
   2. bx-python (>=0.5.0)
   3. numpy (>=1.8.2)
   4. scipy (>=0.15.1) libraries
3. R with the RColorBrewer and ggplot2 (>2.2.1) packages
4. G++ compiler
5. Samtools (>1.1)
6. Unix sort (which support –V option)
   1. Install the GNU core utilities for Mac OS user

## 1.2.1 Comments on different versions of HiC-Pro

There are two different versions that has been used to analyze the data: 2.9.0 and 2.10.0

* Switched from 2.9.0 to 2.10.0 (newer) in **January 2018**
* both versions work fine but I needed to add changes to their scripts for it to run in MacOS – refer to troubleshoot section
* HiC-Pro\_2.9.0 is stored in Macintosh\_HD\_2 in ‘**HiC-Pro\_VNK’** folder
* HiC-Pro\_2.10.0 is stored in **/usr/local/bin (use this)**

## 1.2.2 Installation of the required dependencies

The following should be checked to see if appropriate versions of dependencies are installed:

### miniconda2

‘**miniconda2**’ should be installed, but alternative ‘Anaconda’ is also available

-check with conda list at a Terminal prompt, and then we should be able to see the following list:

conda list

**# packages in environment at /Users/gelab/miniconda2:**

#

asn1crypto                0.22.0 py27\_0

bx                        0.3.0 <pip>

**bx-pytho**n               **0.7.3**                     <pip>

cffi                      1.10.0 py27\_0

conda                     4.3.21 py27\_0

.

.

.

### python libraries: pysam, numpy, scipy, bx

Check to see if all the python libraries are up to date in python 2.7.13:

**Python 2.7.13**

>>> import pysam

>>> pysam.\_\_version\_\_

'0.12.0.1'

>>> import numpy

>>> numpy.\_\_version\_\_

'1.13.3'

>>> import scipy

>>> scipy.\_\_version\_\_

'0.19.1'

>>> import bx

>>> bx.\_\_version\_\_

'0.5.0'

### samtools

Program: **samtools** (Tools for alignments in the SAM format)

Current version: **1.2** (using htslib 1.2.1)

### bowtie 2

**Bowtie 2 current version 2.2.5** by Ben Langmead (langmea@cs.jhu.edu, www.cs.jhu.edu/~langmea)

### unix sort

|  |
| --- |
| which sort  $ /usr/local/opt/coreutils/libexec/gnubin/sort |

If you don’t see the above path for sort, then check the troubleshoot section. Or else, this will cause an error in HiC-Pro since sort-V option is not available

in Mac OS

## 1.2.3 How to update or install necessary dependencies

### numpy:

Updated numpy by:

|  |
| --- |
| sudo easy\_install ‘numpy >=1.8.2’ |

### scipy

Had to install fortran compiler. If having trouble upgrading scipy, it will need more searching online

### bx (bx-python)

bx-python is the one that needs to be installed

There is another way to install bx-python possibly using conda

|  |
| --- |
| conda install -c bcbio bx-python |

Reference: <https://pypi.python.org/pypi/bx-python>

### Installing GNU:

brew install coreutils findutils gun-tar gnu-sed gawk gnutls gnu-indent gnu-getopt

then add it to PATH by typing the following in ‘~/.bash\_profile’

nano ~/.bash\_profile

|  |
| --- |
| export MANPATH="/usr/local/opt/coreutils/libexec/gnuman:$MANPATH"  alias sort=gsort  export PATH="/usr/local/opt/coreutils/libexec/gnubin:$PATH"  export PATH=/usr/local/Cellar/coreutils/8.27/bin/:$PATH |

Lastly, check with which sort

### Homebrew

Here, it requires **Homebrew** to install gnu and other packages:

Follow the instruction in the reference to install Homebrew: <https://brew.sh/>

**Homebrew** is a free and open-source software package management system that simplifies the installation of software on Apple's macOS operating system

# Running HiC-Pro

1. Before running HiC-Pro, ‘configuration file’ should be tailored to the type of data being analyzed

There are two configuration file ready to be used:

* config\_test\_latest\_mm9.txt (for mm9)
* config\_test\_latest.txt (for hg19)

Edit the following parameters if any of these have been moved or replaced:

* BOWTIE2\_IDX\_PATH: /Users/noekimv2/bowtie2\_indicies/bowtie2\_mm9\_indicies/mm9
* REFERENCE\_GENOME: mm9
* GENOME\_SIZE: chrom\_mm9.sizes
* GENOME\_FRAGMENT: MboI\_resfrag\_mm9.bed
* LIGATION SITE: GATCGATC (for MboI restriction enzyme)
* BIN\_SIZE: 100000
* Optional: CAPTURE\_BED: /Users/noekimv2/OID44043\_mm9\_160929\_capture\_targets.bed
  + Used only when analyzing capture HiC
* The rest were left at default

1. Run HiC-Pro:

At a Terminal prompt:

HiC-Pro

usage : HiC-Pro -i INPUT -o OUTPUT -c CONFIG [-s ANALYSIS\_STEP] [-p] [-h] [-v]

e.g.

time /usr/local/bin/HiC-Pro\_2.10.0/bin/HiC-Pro -i CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2/ -o CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2\_output -c config\_test\_latest\_mm9.txt

1. Navigate through output files

Mostly in **hic\_results/** folder

## 2.1 Generate higher resolution matrix

In the configuration file, default bin size is specified. After HiC-Pro pipeline has already analyzed the data, higher resolution can be acquired:

1. In configuration file, BIN\_SIZE should be edited to desired value (e.g. 5000)
2. Run the following command as shown in the example

e.g.

time ~/HiC-Pro\_2.9.0/bin/HiC-Pro -i PLAC-seq\_sample\_HiC-Pro\_results\_2/hic\_results/data/ -o [output directory] –c config\_test\_latest.txt -s build\_contact\_maps -s ice\_norm

Run HiC-Pro 2.9.0

--------------------------------------------

Mon Nov 27 14:35:25 EST 2017

Generate binned matrix files ...

/Users/noekimv2/HiC-Pro\_2.9.0/scripts/build\_raw\_maps.sh -c /Users/noekimv2/config\_test\_latest\_mm9.txt

--------------------------------------------

Mon Nov 27 14:37:41 EST 2017

Run ICE Normalization ...

/Users/noekimv2/HiC-Pro\_2.9.0/scripts/ice\_norm.sh -c /Users/noekimv2/config\_test\_latest\_mm9.txt >> hicpro.log

real 5m53.400s

user 5m15.984s

sys 0m17.662s

# Post-processing HiC-Pro output files

## 3.1 Hicpro2juicebox

The purpose is to help us visualize contact heatmap in ‘Juicebox software’ which is part of the **Juicer** pipeline

With Juicebox software, you can also see fold change between different conditions (control vs. KO)

The following shows how to run hicpro2juicebox.sh:

bash hicpro2juicebox.sh -h

usage : hicpro2juicebox -i VALIDPAIRS -g GSIZE -j JUICEBOXJAR [-r RESFRAG] [-t TEMP] [-o OUT] [-h]

e.g.

**Command**: bash /usr/local/bin/HiC-Pro\_2.10.0/bin/utils/hicpro2juicebox.sh

-g : /usr/local/bin/HiC-Pro\_2.10.0/annotation/chrom\_mm9.sizes

-j : /Users/noekimv2/Other\_softwares/juicebox/out/artifacts/Juicebox\_tools\_jar/juicebox\_tools.jar

-r : /usr/local/bin/HiC-Pro\_2.10.0/annotation/MboI\_resfrag\_mm9.bed

-i:

/Users/noekimv2/CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2\_output/hic\_results/data/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2\_allValidPairs

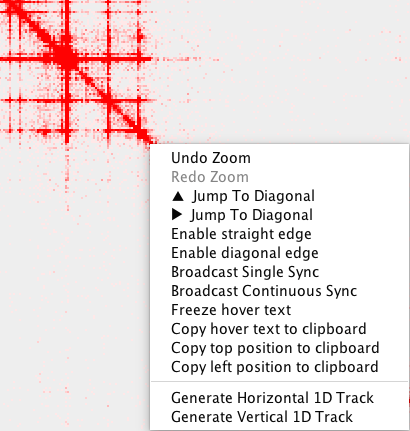
### 3.1.1 Virtual 4C plot



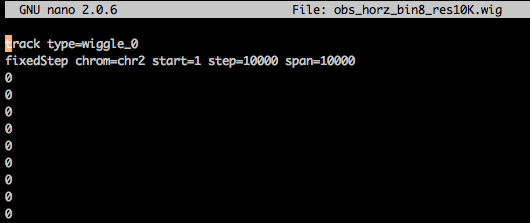
Juicebox is located in Applications

* If user would like to download updated version, check out the reference
* <http://aidenlab.org/software.html>

With Juicebox software, virtual 4C plot can be generated by right clicking on the anchor of interest, then select ‘Generate Horizontal 1D Track’



* Then, this generates ‘wiggle’ file in ‘juicebox’ folder which can be uploaded to UCSC genome browser
* Add ‘track type=wiggle\_0’ as the first line of the generated .wig file



* Or, it can be converted to ‘bigwig’ format using ‘wigToBigwig’ tool, then upload it to WashU EpiGenome Browser
  + .bigwig file should be uploaded to Amazon S3 or other web services that can generate URL links
  + WashU epigenome browser accepts URL for .bigwig files
* Alternatively, there is another method that can be visualized in 3D Genome Browser, and the general steps are described in section 3.5.

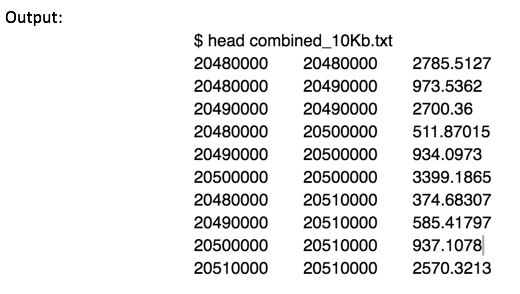
### 3.1.2 Juicebox dump function

Alternative method to generate virtual 4c plot

**\*\*Juicebox\_tools.jar location**: **smb://shares.dkisilon.niddk.nih.gov/dkgelab**/Mac 1/Victoria/Software/juicebox/out/artifacts/Juicebox\_jar/

This following **command** will dump the observed matrix of chromosome 1 from 20.48MB to 40.96MB with KR normalization at 10000 BP resolution

sudo java -jar **juicebox\_tools.jar dump** observed KR combined.hic 1:20480000:40960000 1:20480000:40960000 BP 10000 combined\_10Kb.txt



## 3.2 hicpro2higlass

The purpose is to visualize contact heatmap in HiGlass visualization tool that can be used with HiPiler

bash hicpro2higlass.sh

usage : hicpro2higlass -i INPUT -r RESOLUTION -c CHROMSIZE [-n] [-h]

Use option -h|--help for more information

## 3.3 hicpro2fithic

The purpose of hicpro2fithic is to generate input file for fit-hi-c and to analyze the significant interactions based on all valid read pairs from HiC-Pro

python hicpro2fithic.py

usage: hicpro2fithic.py [-h] -i MATRIX -b BED [-s BIAS] [-o OUTPUT]

[-r RESOLUTION]

e.g.

command:

time python ~/HiC-Pro\_2.9.0/bin/utils/hicpro2fithic.py

**-b:**

/Users/noekimv2/CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2\_output\_res1k\_5k\_10k/hic\_results/matrix/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2/raw/1000/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2\_1000\_abs.bed

**-i:**

/Users/noekimv2/CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2\_output\_res1k\_5k\_10k/hic\_results/matrix/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2/iced/1000/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2\_1000\_iced.matrix

**-s:**

/Users/noekimv2/CHIC\_AdipoPE\_Brd4\_Cre\_GFP\_D2\_output\_res1k\_5k\_10k/hic\_results/matrix/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2/iced/1000/Sample\_CHiC\_AdipoPE\_Brd4\_GFP\_D2\_1000\_iced.matrix.biases

$ time python /usr/local/bin/HiC-Pro\_2.10.0/bin/utils/hicpro2fithic.py -b /Users/noekimv2/hichip\_naive\_b2t2\_fastq\_output\_res80k\_90k/hic\_results/matrix/hichip\_naive\_B2T2/raw/90000/hichip\_naive\_B2T2\_90000\_abs.bed -i /Users/noekimv2/hichip\_naive\_b2t2\_fastq\_output\_res80k\_90k/hic\_results/matrix/hichip\_naive\_B2T2/iced/90000/hichip\_naive\_B2T2\_90000\_iced.matrix -s /Users/noekimv2/hichip\_naive\_b2t2\_fastq\_output\_res80k\_90k/hic\_results/matrix/hichip\_naive\_B2T2/iced/90000/hichip\_naive\_B2T2\_90000\_iced.matrix.biases

|  |
| --- |
| **Output files** |
| fithic.biases.gz |
| fithic.fragmentMappability.gz |
| fithic.interactionCounts.gz |

### 3.3.1 Fit-Hi-C: Call significant loops

Step 1: Generate input files of Fit-Hi-C using hicpro2fithic.py

* Hicpro2fithic.py is located in ‘PATH/TO/HiC-Pro/bin/utils’



Step 2: Open RStudio

* Located in Applications

Step 3: Do the following in RStudio with appropriate files

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

## Set working directory where it contains output of hicpro2fithic.py

setwd("~/FitHiC\_files")

## Define ‘intersfile’

intersfile <- file.path("~/FitHiC\_files/sample\_PLAC\_PolII\_GFP\_D2\_40000/fithic.interactionCounts.gz")

## Define ‘fragsfile’

fragsfile <-

file.path("~/FitHiC\_files/sample\_PLAC\_PolII\_GFP\_D2\_40000/fithic.fragmentMappability.gz")

## Define ‘output directory’

outdir <- file.path("~/FitHiC\_files/output\_GFP\_40000")

## Define ‘biasfile’

biasfile <- file.path("~/FitHiC\_files/sample\_PLAC\_PolII\_GFP\_D2\_40000/fithic.biases.gz")

## Call FitHiC

library(FitHiC)

## Run FitHiC with defined parameters

FitHiC(fragsfile,intersfile,outdir,biasfile, libname="test",noOfBins=200,visual=TRUE)

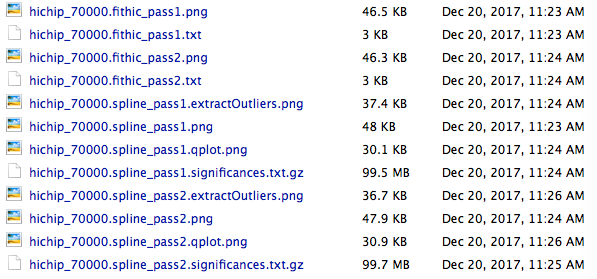
\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

Important note:

* Must be in the same order
* ‘libname’ can be any label that can be identifiable by user

Step 4: Check if the output files are generated in output directory

* The following should be in the output folder
* [libname].spline\_pass2.significances.txt.gz is used to visualize loops in WashU epigenome browser after converting to the appropriate format (pairwise interaction format)



## 3.4 Hipro\_macs2

The purpose of hicpro\_macs2 is to generate an input file for WashU EpiGenome Browser to visualize 1D HiChIP (ChIP peaks)

bash hicpro\_macs2.sh

To visualize 1D HiChIP the following .bigwig is needed:

~/HiC-Pro\_output/hic\_results/data/dixon\_2M/macs2\_output/**test\_DE+SC\_macs2input\_1Dtrack\_treat\_sorted.bigwig**

Note: The highlighted part should have different names since it is user defined and sample names are different

Note for Victoria:

* For now, this works for two samples (usually two samples per HiC-Pro run)
* Finalize this with real data
* The output file is ‘bigwig’ format which can be uploaded to Amazon 3S or Cyverse, use its URL to visualize 1D HiChIP track to WashU Epigenome Browser
* Latest one: hicpro\_macs2\_180426.sh

## 3.5 Virtual 4C plot for 3D genome browser

If the user wishes to use 3D genome browser (Penn State), then the interaction files need to be converted to .btr format as seen in the following summary

References:

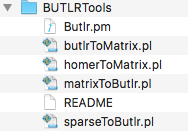
3D genome browser: <http://promoter.bx.psu.edu/hi-c/>

* Web browser

Homer: <http://homer.ucsd.edu/homer/interactions/HiCinteractions.html>

* Installed in local station

BUTLRtools: <http://promoter.bx.psu.edu/hi-c/butlr.html>

* Installed in local station



However, many issues/limitations arise while following these steps including:

* Resolutions 🡪 this is a big issue because homer cannot generate matrix size bigger than 9.00e+08 which is required as an input of BUTLR tools
* Incongruent matrix size 🡪 this is addressed in the reference

## 3.6 Hichipper: specific to analyzing HiChIP data

* This pipeline takes HiC-Pro output files for downstream analysis
* This has not been explored yet
* Installed in ~/Other\_softwares/hichipper

Reference: <https://www.biorxiv.org/content/early/2017/09/21/192302>

# Troubleshoot

## Comments

### 4.1 #1 Issue: HiC-Pro runs into an error while running —make:\*\*\*[bowtie\_pairing] Error 1

Tue Oct  3 18:12:46 EDT 2017

Pairing of R1 and R2 tags ...

/software/HiC-Pro\_2.9.0/scripts/bowtie\_pairing.sh -c /Volumes/Macintosh\_HD/software/config\_test\_latest\_mm9.txt >> hicpro.log

**make: \*\*\* [bowtie\_pairing] Error 1**

Log from mergeSAM.log showed that:

## Merging forward and reverse tags ...

**Forward and reverse reads not paired. Check that BAM files have the same read names and are sorted.**

**Solution**:

After troubleshooting, the following helped fixing this particular HiC-Pro error:

1. Update **samtools**
   1. In my case: from version 1.2 🡪 1.5
2. Change the script from HiC-Pro\_2.X.0/scripts/bowtie\_combine.sh
   1. Line 50: … -T ${TMP\_DIR} … 🡪 -T ${HOME}
   2. This change was made because an error with ‘mapped\_2hic\_fragments.sh’ showed that TMP\_DIR did not exist

### 4.2 #2 Issue: sort: stray character in field spec: invalid field specification ‘2,2V’

**Solution**:

When running HiC-Pro pipeline and you see this error, the following should be double checked:

1. check and see if GNU utility is correctly installed and have been added to the path

*Step 1*: $ which sort

*Step 2*: if you do not see $/usr/local/opt/coreutils/libexec/gnubin/sort,

* 1. then type source ~/.bash\_profile
  2. check again by which sort

1. Make sure the following is not commented out in ‘.bash\_profile’

|  |
| --- |
| export MANPATH="/usr/local/opt/coreutils/libexec/gnuman:$MANPATH"  alias sort=gsort  export PATH="/usr/local/opt/coreutils/libexec/gnubin:$PATH"  export PATH=/usr/local/Cellar/coreutils/8.27/bin/:$PATH |

### 4.3 #3 Issue: Interference with iTerm color scheme

If you are using iTerm, iTerm color scheme somehow is interfered when the following was added to ‘~/.bash\_profile’

##GNU utility

export MANPATH="/usr/local/opt/coreutils/libexec/gnuman:$MANPATH"

alias sort=gsort

**export PATH="/usr/local/opt/coreutils/libexec/gnubin:$PATH"**

export PATH=/usr/local/Cellar/coreutils/8.27/bin/:$PATH

|  |  |
| --- | --- |
| what we should see | without ANSI color |

**Solution:**

When the blue text is commented out in ‘~/.bash\_profile’, the color ANSI comes back to normal

You can comment the line out when not using HiC-Pro