

Workshop 10: Hi-C

Fides Lay

QCB Fellow

lay.fides@gmail.com

Workshop 10 Outline

Day 1:

Introduction to the 3D genome

Overview of Hi-C experiments: wet lab and computational consideration

Quick review of linux, Hoffman2 and high-throughput sequencing glossary.

Aligning Hi-C library using HiC-Pro

Day 2:

Aligning Hi-C library using HiC-Pro, continued.

Visualization of data using HiC-Plotter

Day 3:

Identification of significant chromatin interactions using GOTHIC.

Calculate directionality index using HiTC.

Overview of other useful Hi-C tools: hicup, HOMER

Day 1

Higher-Order Chromatin Structures

Nucleosomal scale

1 pb - ~10 kb

EPIGENETIC
MODIFICATIONS

NUCLEOSOMES

Supranucleosomal scale

~10kb - ~800kb

INTRA-TAD
DYNAMICS

TOPOLOGICALLY
ASSOCIATING
DOMAINS = TADs

~3Mb

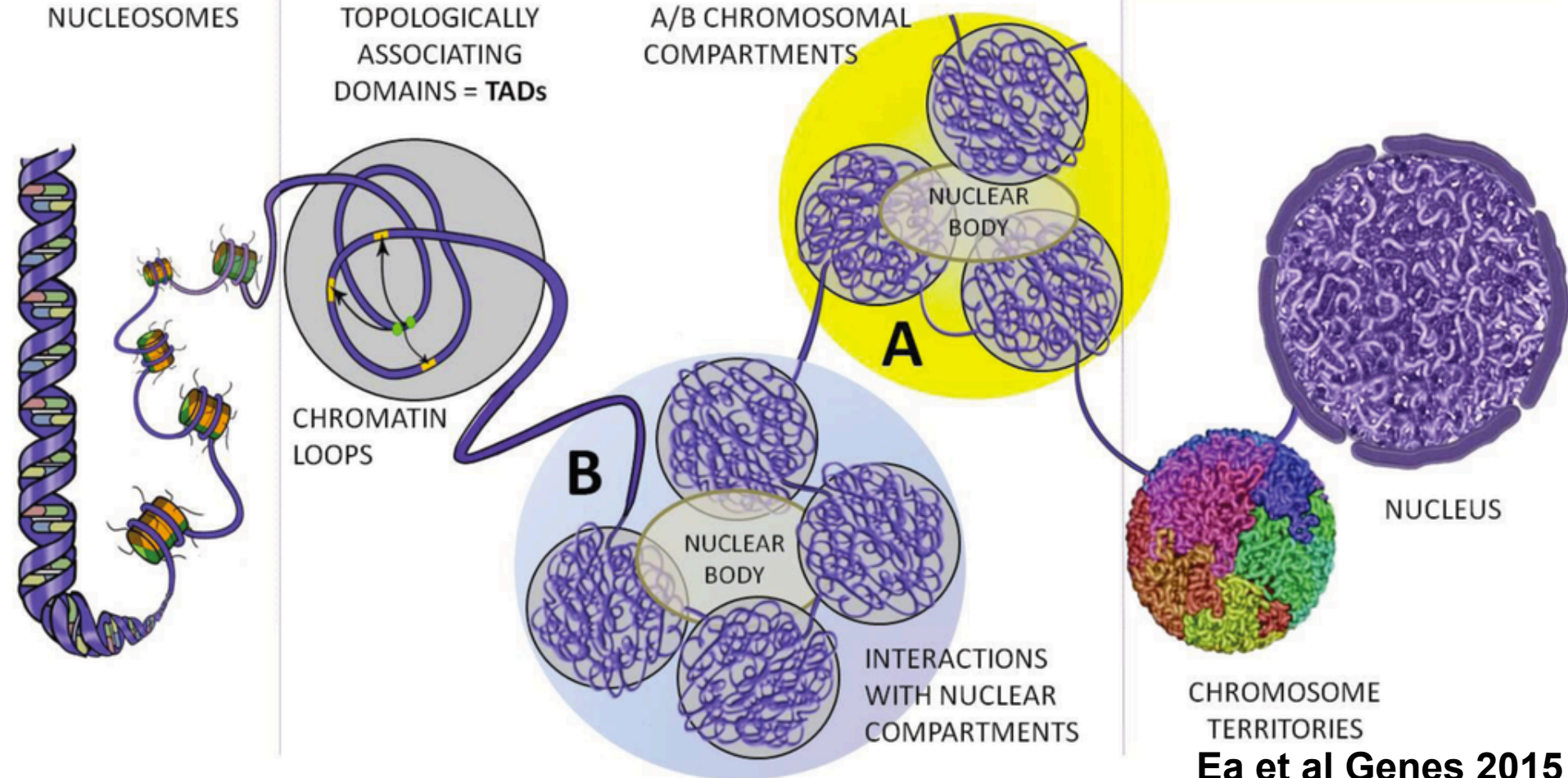
INTER-TAD
DYNAMICS

A/B CHROMOSOMAL
COMPARTMENTS

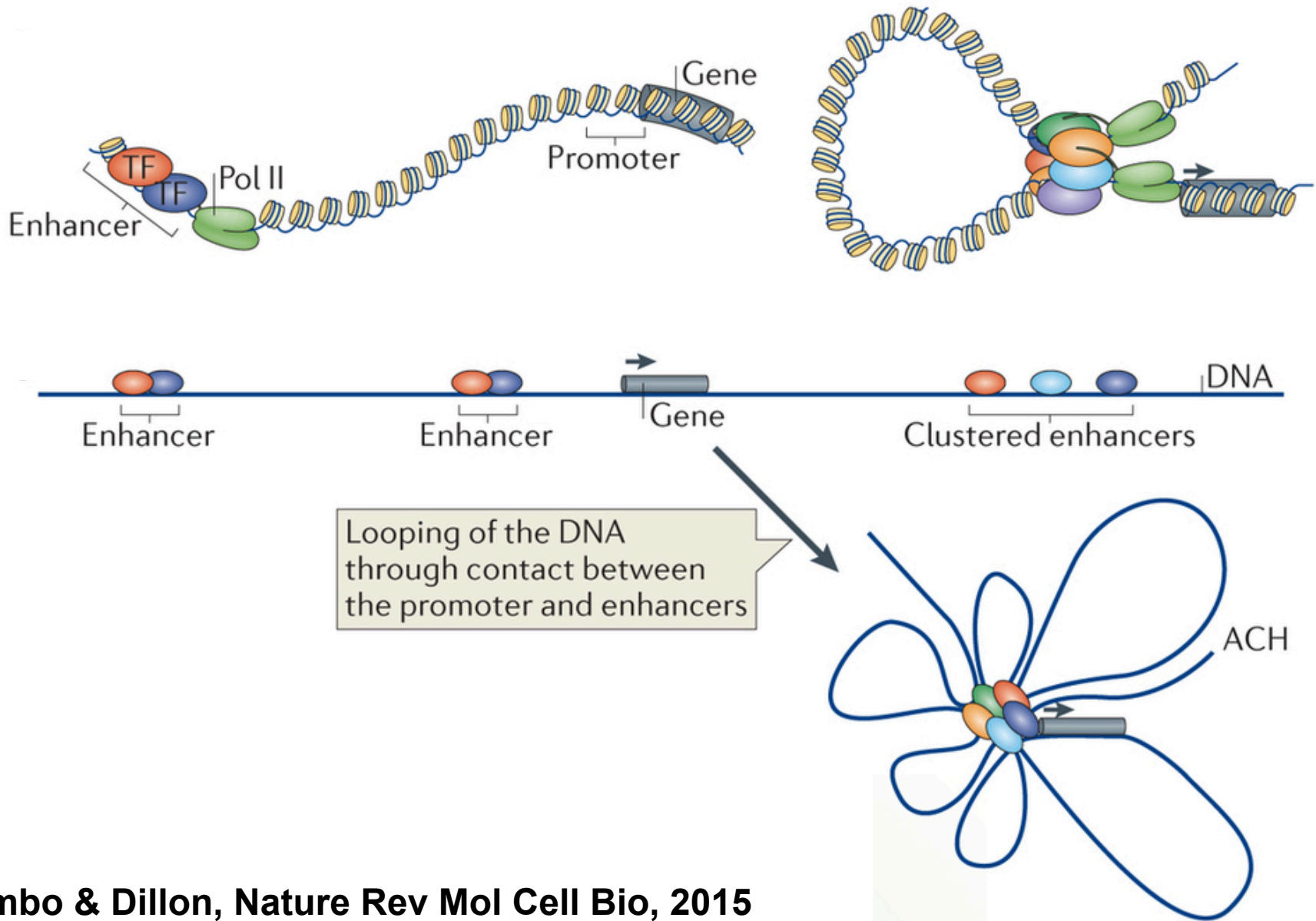
Nuclear scale

~100Mb - ~3000Mb

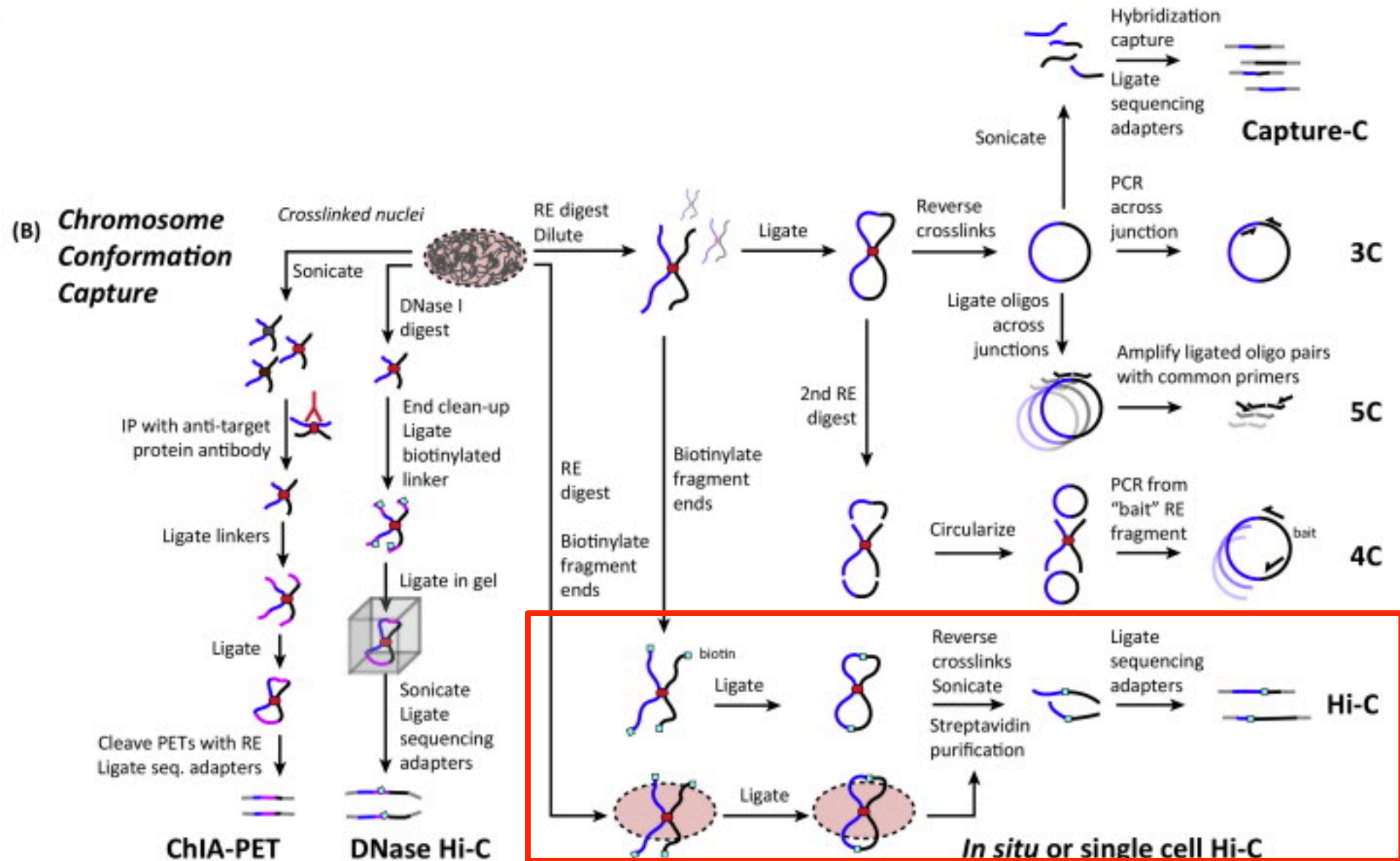
NUCLEAR
POSITIONING



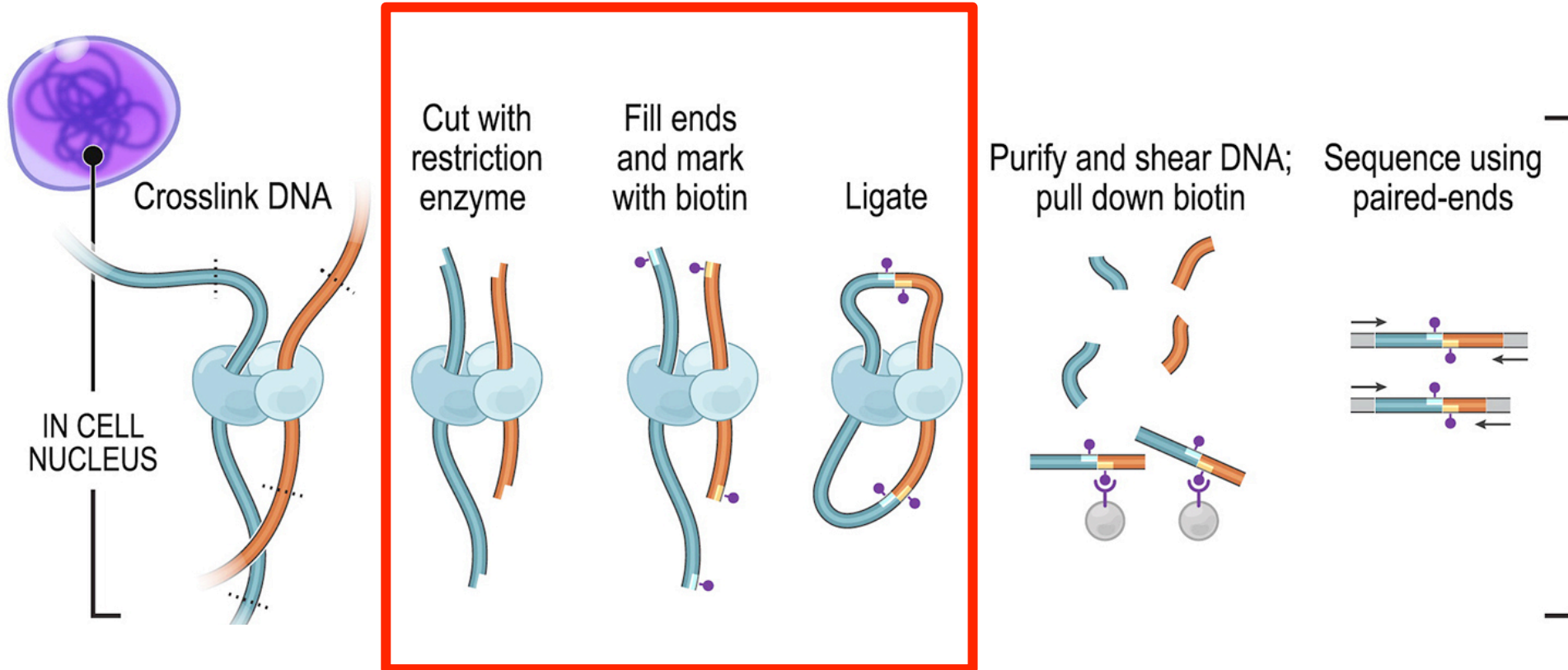
Enhancer-Promoter Looping



Chromatin Interaction Assays



Hi-C Protocol Overview



Ligation Junction

5'...A[∨]AGCT T...3'
3'...T TCGA_∧A...5'



5'...A[∨]**AGCT** ...3'
3'...T TCGA_∧...5'

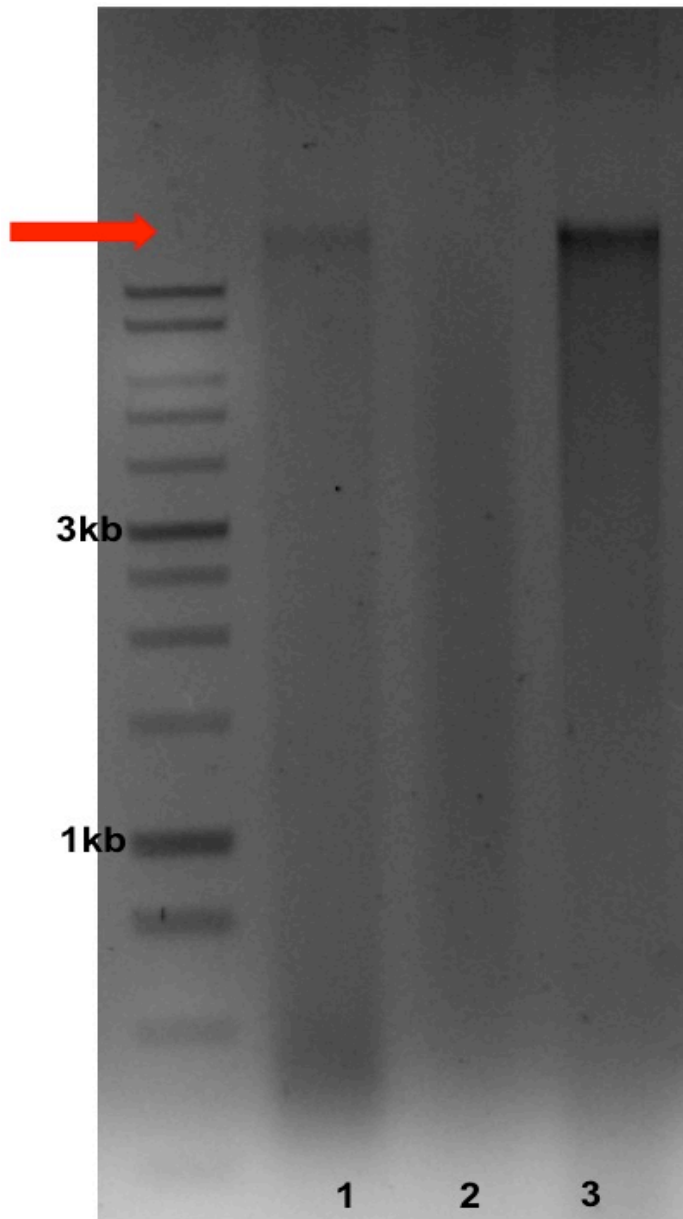


5'...[∨]AGCT T...3'
3'...**TCGA**_∧A...5'



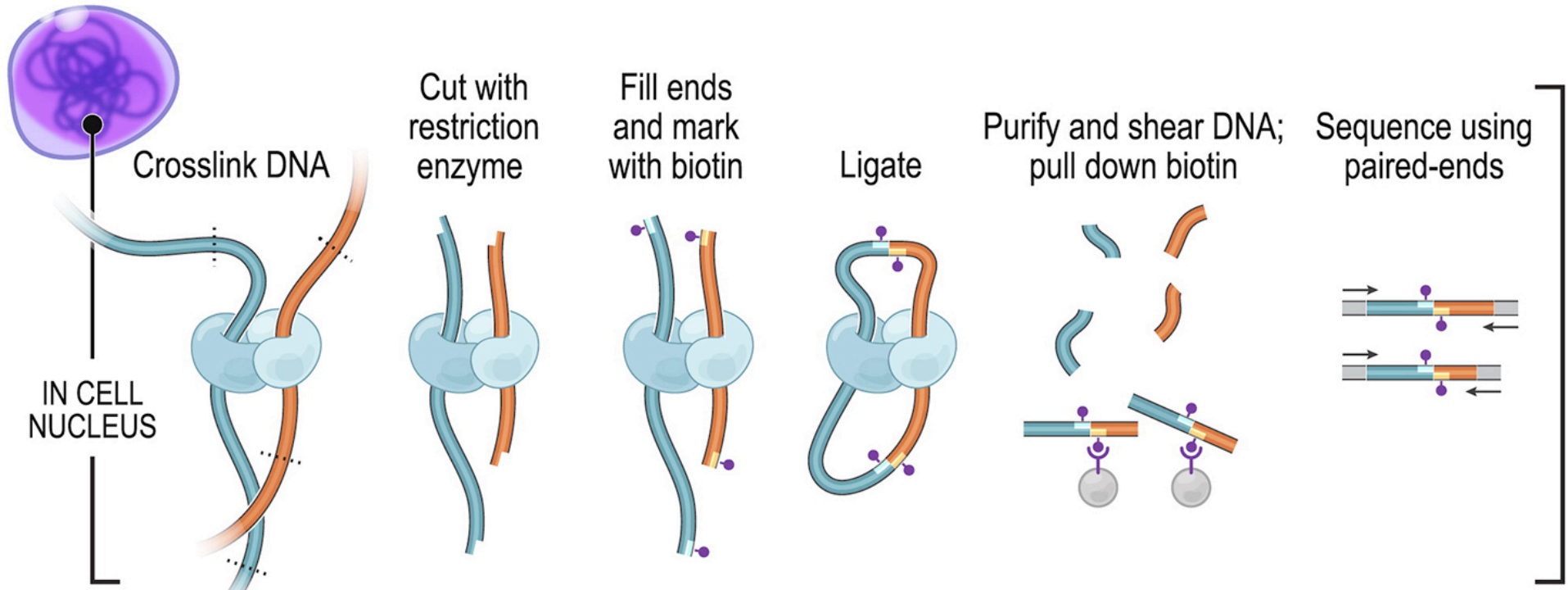
5'...A**AGCT**AGCTT...3'
3'...TTCGAT**TCGA**A...5'

Hi-C Quality Control



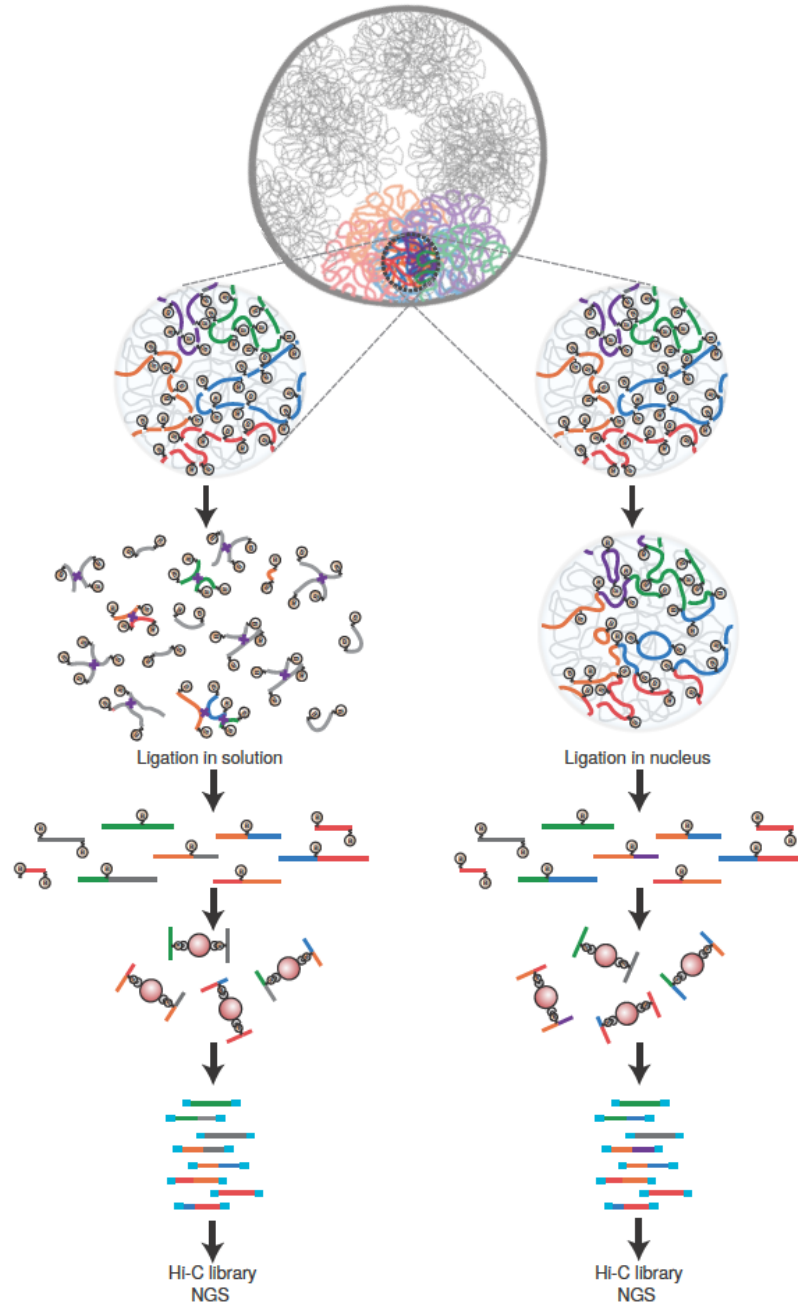
Lane 1: Undigested Templated
Lane 2: Digestion Products
Lane 3: Ligation Products

Hi-C Protocol Overview



Hi-C Protocol Overview

**Dilution
(Lieberman-
Aiden et al
2009)**



**In-Situ
(Rao et al
2014)**

Nagano et al 2015

Hi-C Protocol Overview

	Dilution	In-Situ
# Cells	1-2.5x10 ⁷	2-5x10 ⁶
Fixative	1% Formaldehyde	1%Formaldehyde
Restriction Enzyme	6-Cutter (HindIII)	4-Cutter (Mbol)
Digestion Time	O/N	2hrs-O/N
Solubilization	SDS, 65C	N/A
Ligation Time	4hrs-O/N	4hrs
Sequencing Method	50PE (Lieberman-Aiden et al 2009)	100PE (Rao et al, 2014)
Coverage Needed	~1 billion paired-reads	~300 million paired-reads

RESEARCH

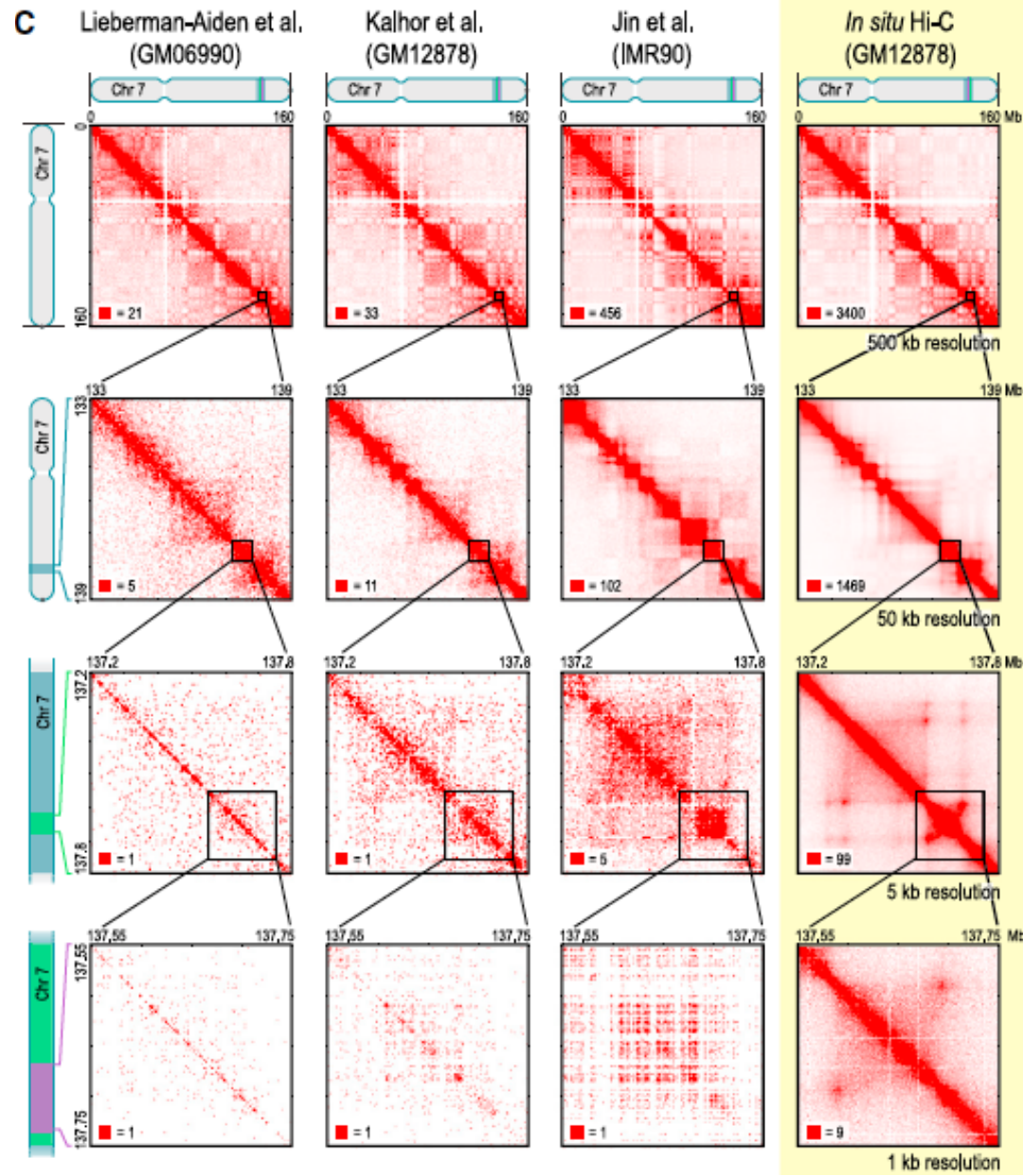
Open Access

Comparison of Hi-C results using in-solution versus in-nucleus ligation



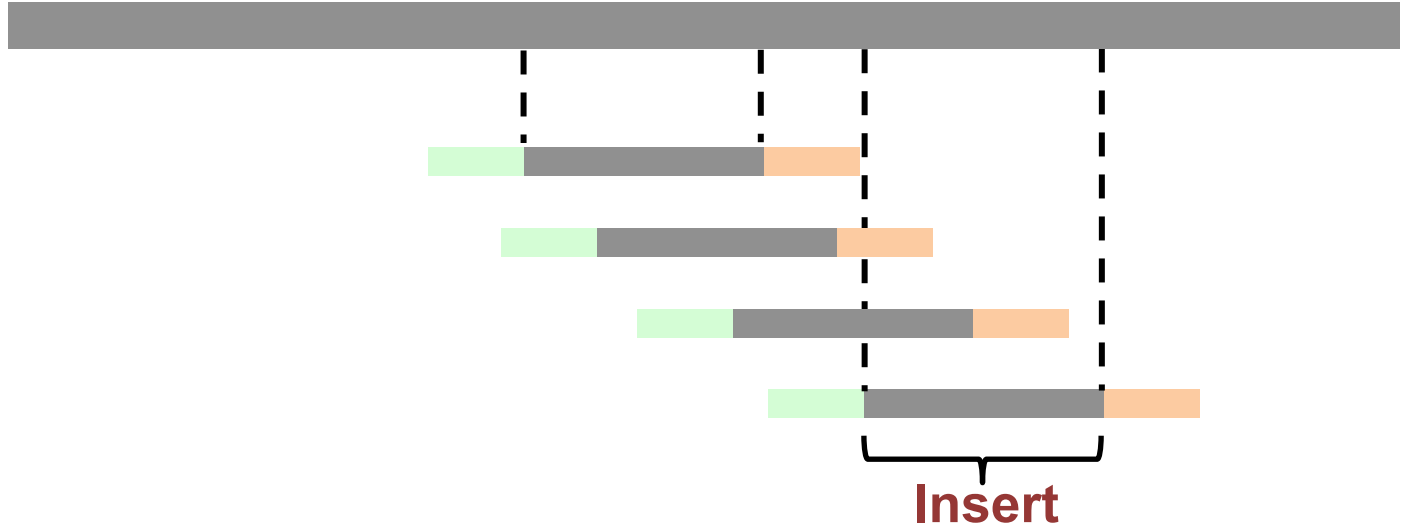
Takashi Nagano^{*†}, Csilla Várnai[†], Stefan Schoenfelder, Biola-Maria Javierre, Steven W. Wingett and Peter Fraser^{*}

In-Situ Hi-C Generates Higher Resolution Interaction Data

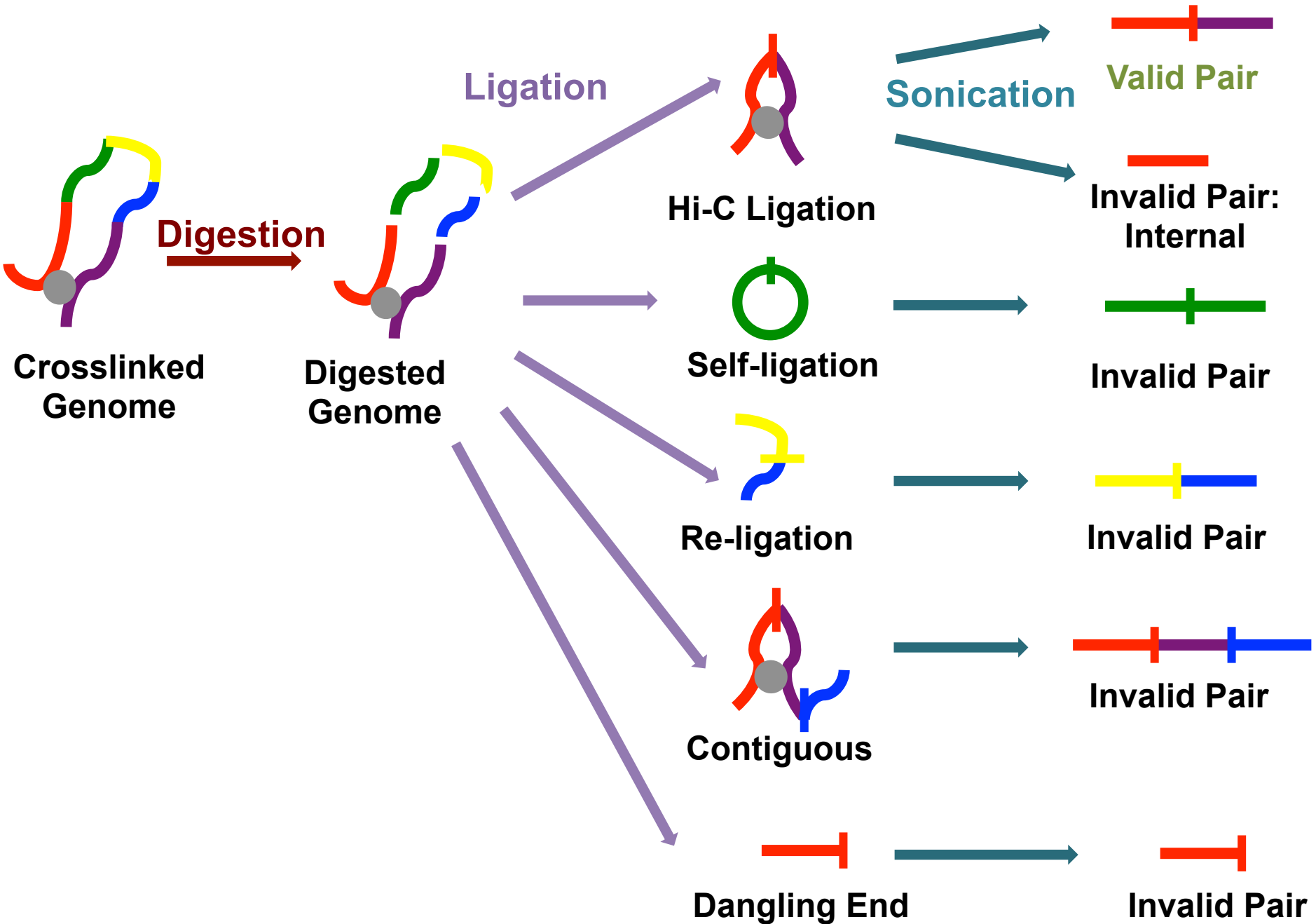


Paired-end Alignment

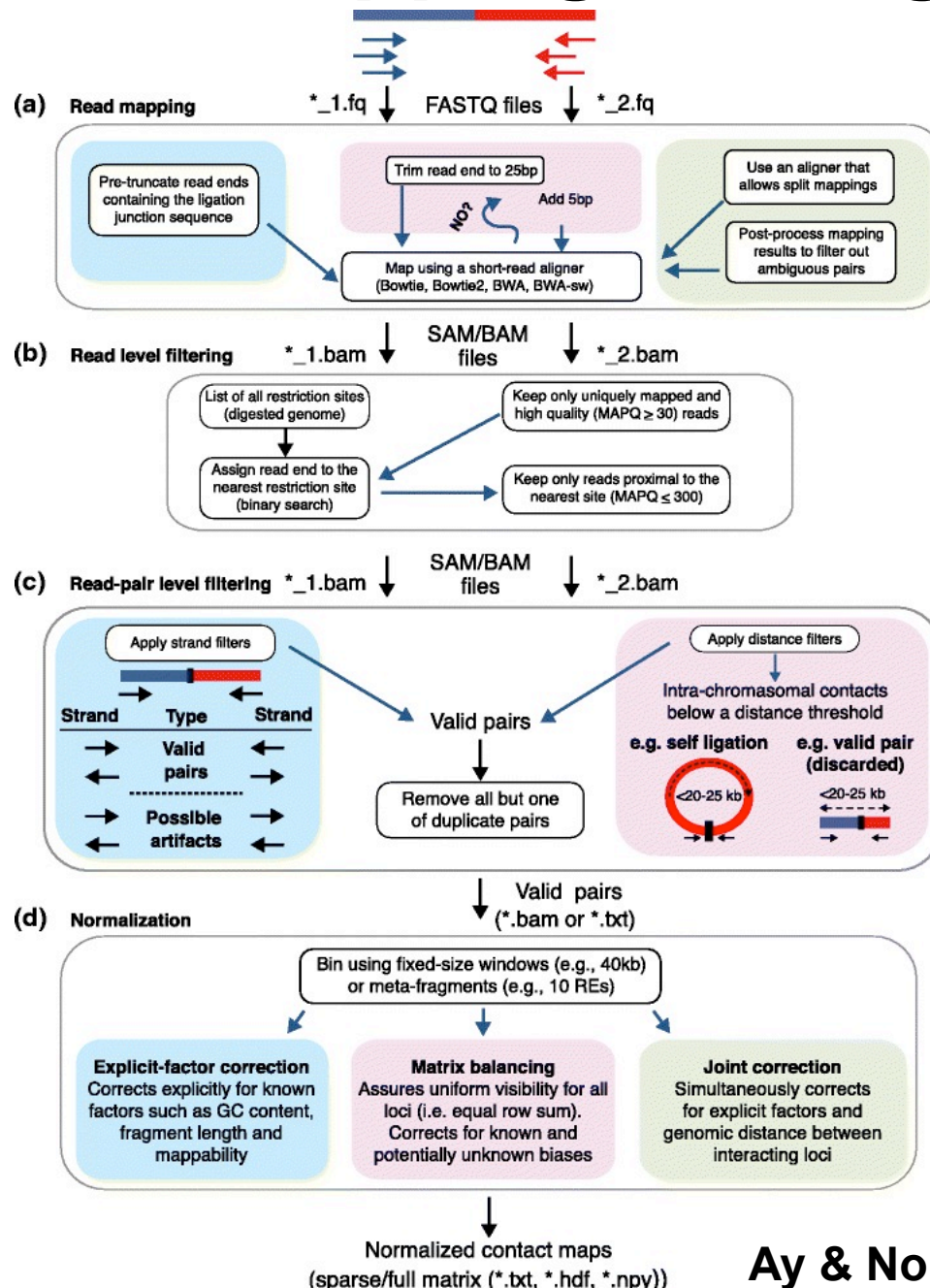
Reference
Genome



Common Artifacts in Hi-C Mapping



Hi-C Mapping Strategy



Source of Biases

- Fragment Length
- Digestion and Ligation efficiency
- Mappability
- Coverage
- GC Content
- Genomic Distance
- Restriction site density
- Chromatin state: open/closed

Iterative Correction (Matrix Balancing)

NATURE METHODS | ARTICLE



Iterative correction of Hi-C data reveals hallmarks of chromosome organization

Maxim Imakaev, Geoffrey Fudenberg, Rachel Patton McCord, Natalia Naumova, Anton Goloborodko, Bryan R Lajoie, Job Dekker & Leonid A Mirny

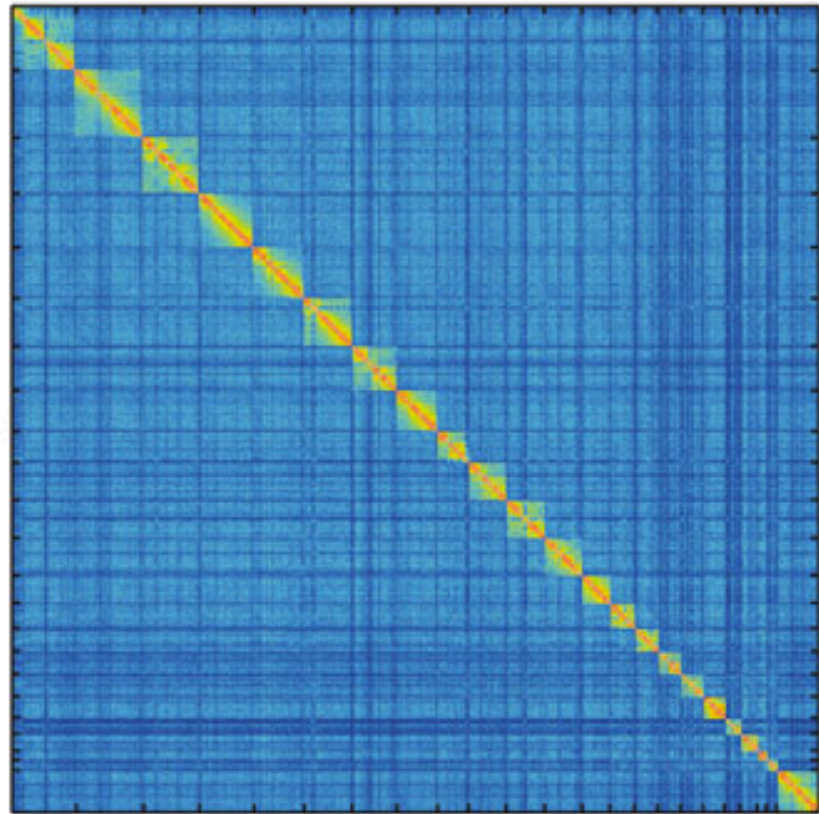
[Affiliations](#) | [Contributions](#) | [Corresponding authors](#)

Nature Methods **9**, 999–1003 (2012) | doi:10.1038/nmeth.2148

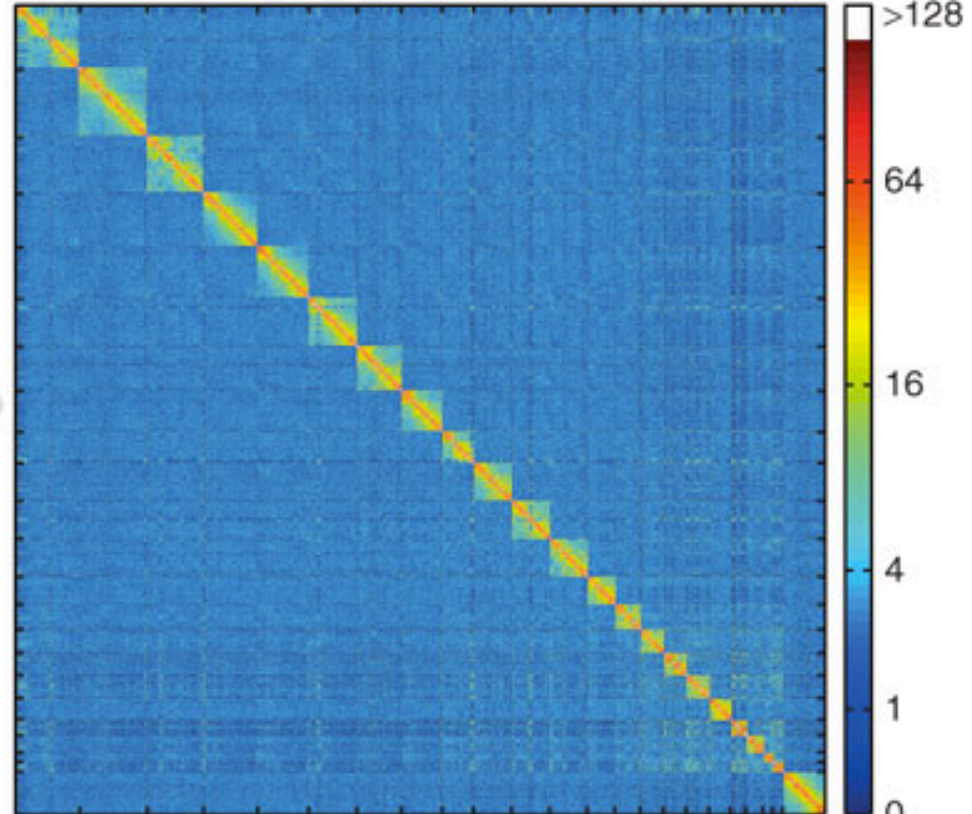
Received 14 May 2012 | Accepted 04 August 2012 | Published online 02 September 2012

Iterative Correction (Matrix Balancing)

Raw



Iteratively corrected



Software Tools for Hi-C Analysis

Tool	Short-read aligner(s)	Mapping improvement	Read filtering	Read-pair filtering	Normalization	Visualization	Confidence estimation	Implementation language(s)
HiCUP [46]	Bowtie/Bowtie2	Pre-truncation	✓	✓	–	–	–	Perl, R
Hiclib [47]	Bowtie2	Iterative	✓ ^a	✓	Matrix balancing	✓	–	Python
HiC-inspector [131]	Bowtie	–	✓	✓	–	✓	–	Perl, R
HIPPIE [132]	STAR	✓ ^b	✓	✓	–	–	–	Python, Perl, R
HiC-Box [133]	Bowtie2	–	✓	✓	Matrix balancing	✓	–	Python
HiCdat [122]	Subread	– ^c	✓	✓	Three options ^d	✓	–	C++, R
HiC-Pro [134]	Bowtie2	Trimming	✓	✓	Matrix balancing	–	–	Python, R
TADbit [120]	GEM	Iterative	✓	✓	Matrix balancing	✓	–	Python
HOMER [62]	–	–	✓	✓	Two options ^e	✓	✓	Perl, R, Java
Hicpipe [54]	–	–	–	–	Explicit-factor	–	–	Perl, R, C++
HiBrowse [69]	–	–	–	–	–	✓	✓	Web-based
Hi-Corrector [57]	–	–	–	–	Matrix balancing	–	–	ANSI C
GOTHic [135]	–	–	✓	✓	–	–	✓	R
HiTC [121]	–	–	–	–	Two options ^f	✓	✓	R
chromoR [59]	–	–	–	–	Variance stabilization	–	–	R
HiFive [136]	–	–	✓	✓	Three options ^g	✓	–	Python
Fit-Hi-C [20]	–	–	–	–	–	✓	✓	Python

SOFTWARE

Open Access



HiC-Pro: an optimized and flexible pipeline for Hi-C data processing

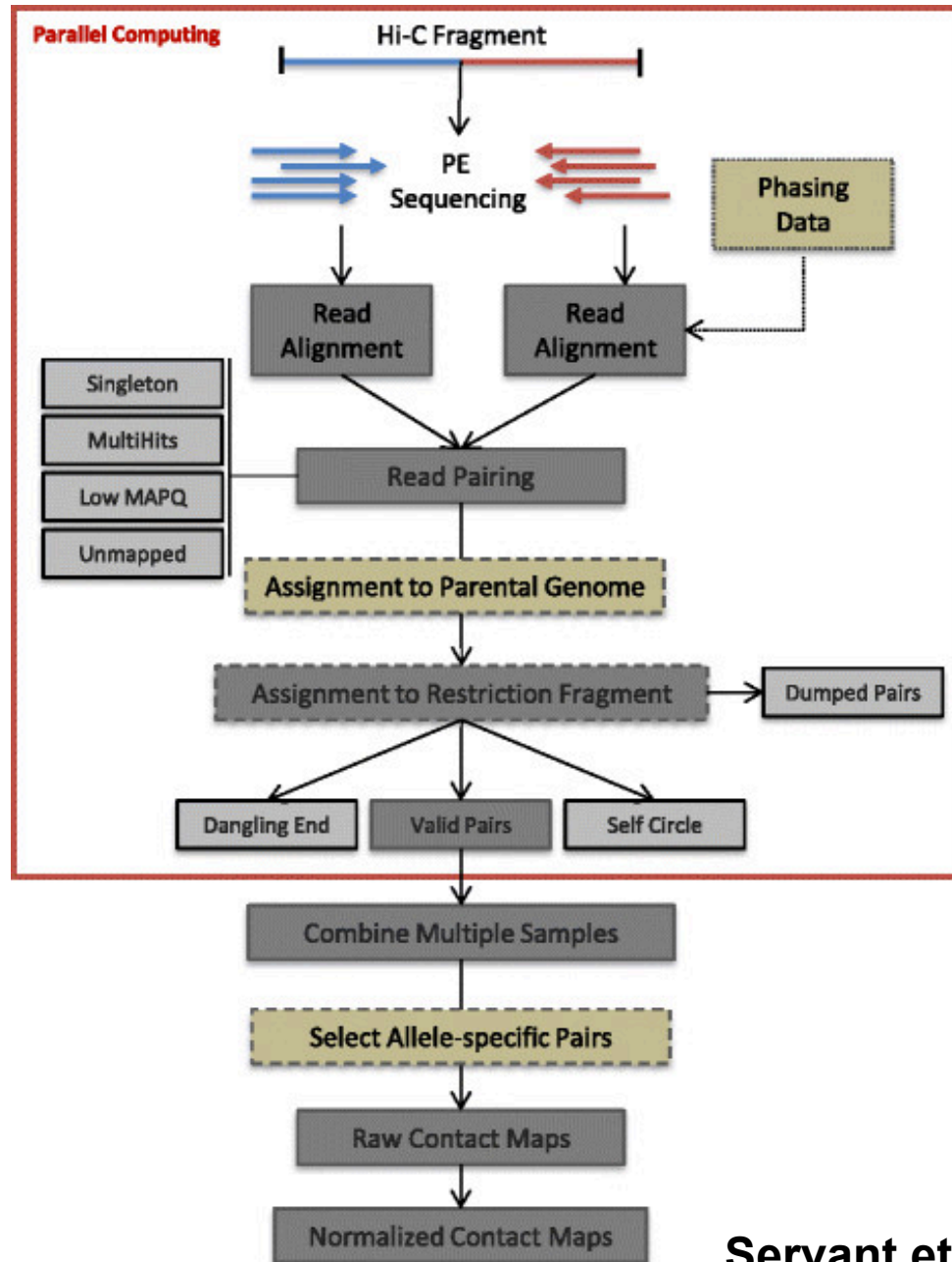
Nicolas Servant^{1,2,3*}, Nelle Varoquaux^{1,2,3}, Bryan R. Lajoie⁴, Eric Viara⁵, Chong-Jian Chen^{1,2,3,6,7,8}, Jean-Philippe Vert^{1,2,3}, Edith Heard^{1,6,7}, Job Dekker⁹ and Emmanuel Barillot^{1,2,3}

Abstract

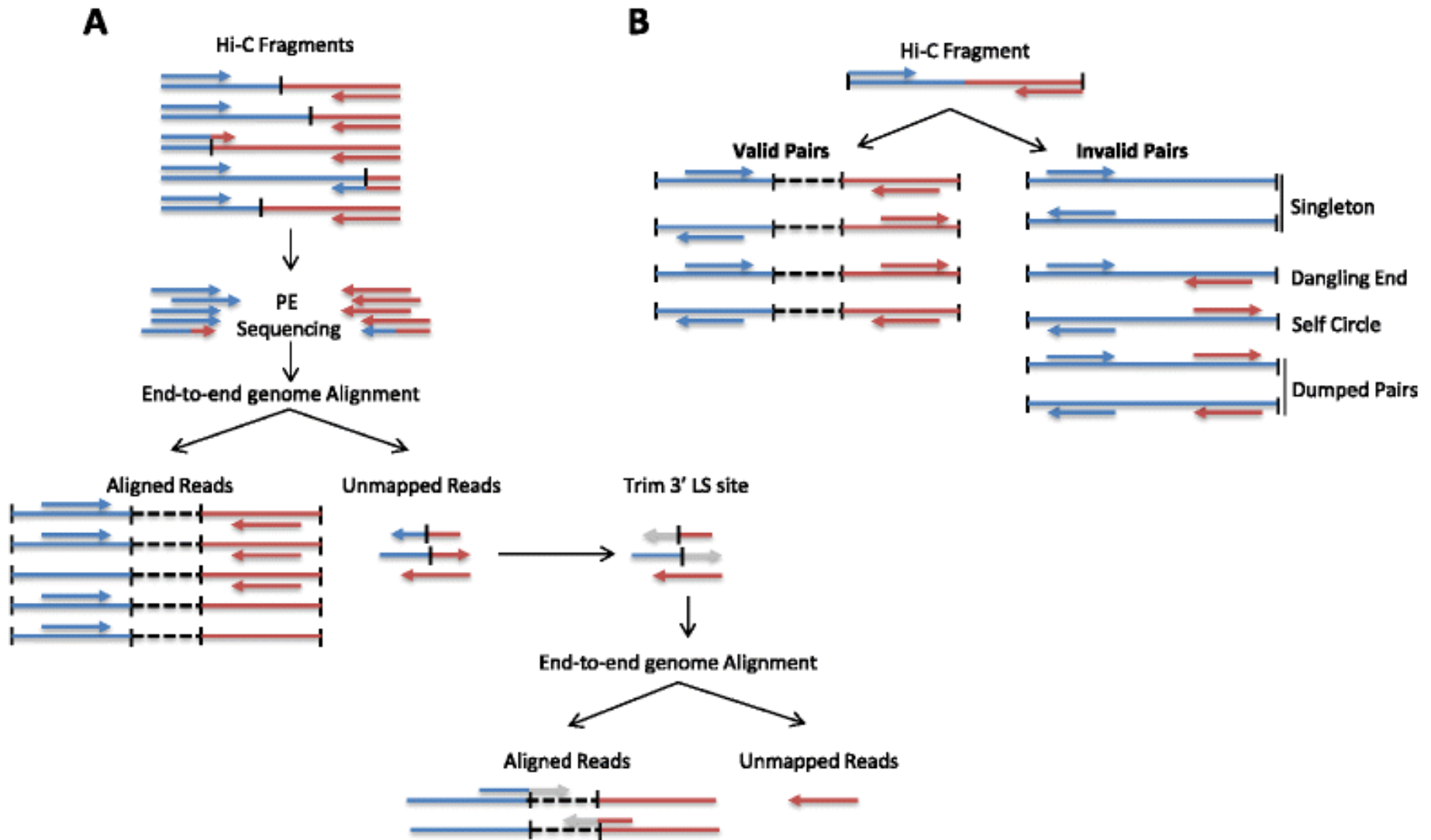
HiC-Pro is an optimized and flexible pipeline for processing Hi-C data from raw reads to normalized contact maps. HiC-Pro maps reads, detects valid ligation products, performs quality controls and generates intra- and inter-chromosomal contact maps. It includes a fast implementation of the iterative correction method and is based on a memory-efficient data format for Hi-C contact maps. In addition, HiC-Pro can use phased genotype data to build allele-specific contact maps. We applied HiC-Pro to different Hi-C datasets, demonstrating its ability to easily process large data in a reasonable time. Source code and documentation are available at <http://github.com/nservant/HiC-Pro>.

Keywords: Chromosome conformation, Hi-C, Bioinformatics pipeline, Normalization

HiC-Pro Workflow



HiC-Pro Mapping and Filtering



Useful UNIX Commands: Covered in Workshop 1 & 2

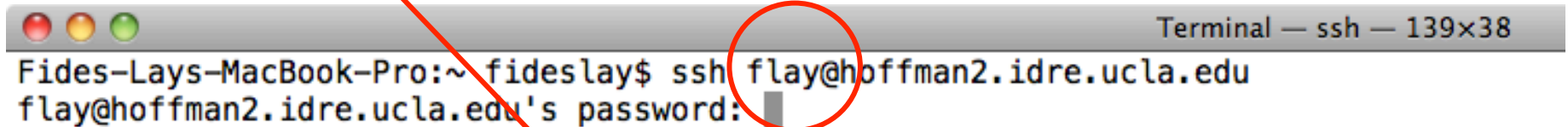
- Where am I? `pwd`
- Change directory `cd        cd ~/data`
- Move up one level `cd ..`
- List files in folder `ls`
- Look at a file `less fileName`
- Copy a file `cp ~/data/file ~/otherdir/`
- Delete a file `rm fileName`
- Delete a directory `rmdir ~/dirName/`
- Move a file `mv ~/data/file ~/otherdir/file`
- Secure copy `scp user@host1:dir/file user@host2:dir/file`
- Compress a file `gzip -c file > file.gz`
- Uncompress a file `gunzip file.gz`
- Make a new folder `mkdir data2`
- Current directory `./`
- Home directory `~/`
- Count lines in a file `wc -l fileName`

Questions/In Doubt/Lost?

- Unix manual: `man functionX`
- Google is your best friend!
- Any of your friendly QCB fellows

Logging into Hoffman2

ssh flay@hoffman2.idre.ucla.edu



A screenshot of a macOS Terminal window. The title bar at the top reads "Terminal — ssh — 139x38". The terminal content shows the command prompt "Fides-Lays-MacBook-Pro:~ fideslay\$" followed by the command "ssh flay@hoffman2.idre.ucla.edu". Below this, it says "flay@hoffman2.idre.ucla.edu's password:" followed by a single character and a cursor. Two red circles are drawn on the image: one around "flay" in the command above the terminal, and another around "flay" in the command inside the terminal. Two red arrows originate from these circles and point towards the text "change to your user ID" at the bottom of the slide.

```
Fides-Lays-MacBook-Pro:~ fideslay$ ssh flay@hoffman2.idre.ucla.edu  
flay@hoffman2.idre.ucla.edu's password: █
```

change to your user ID

Logging into Hoffman2

flay@login3:~ — ssh — 139×38

```
Fides-Lays-MacBook-Pro:~ fideslay$ ssh flay@hoffman2.idre.ucla.edu
flay@hoffman2.idre.ucla.edu's password:
Last login: Wed Mar  2 10:03:31 2016 from 164.67.160.102
Welcome to the Hoffman2 Cluster!
```

```
Hoffman2 Home Page:      http://www.hoffman2.idre.ucla.edu
Consulting:              https://support.idre.ucla.edu/helpdesk
```

All login nodes should be accessed via "hoffman2.idre.ucla.edu".

Please do NOT compute on the login nodes.

Processes running on the login nodes which seriously degrade others' use of the system may be terminated without warning. Use qssh to obtain an interactive shell on a compute node for CPU or I/O intensive tasks.

The following news items are currently posted:

- Globus access changes February 13, 2016
- Mathematica upgraded to 10.3
- IDRE Winter 2016 HPC Classes
- News Archive On Web Site

Enter shownews to read the full text of a news item.

```
[flay@login3 ~]$
```

Useful Tools on Hoffman2

module available

```
[flay@login3 ~]$ module available
```

```
----- /u/local/Modules/modulefiles -----
ATS
R/2.12.0
R/2.12.1
R/2.12.2
R/2.13.2
R/2.15.1
R/2.9.1
R/3.0.1(default)
R/3.1.1
R/3.2.1
R/3.2.3
Rstudio/0.98(default)
Rstudio/R-2.15.1
abacus/6.11
abacus/6.12
abacus/6.13
abacus/6.14(default)
activeperl/5.16(default)
affymetrix/rapt-1.14.3(default)
amber/14
amber-intel11.1-mvapich1.9a2/12
annovar/v2011jun18
annovar/v2011oct02
annovar/v2014jul14(default)
ansys/14.0
armadillo/4.600.4
bamtools/2.1.0(default)
bcftools/1.2
bedtools/2.13.3
bedtools/2.17.0(default)
bedtools/2.23.0
blat/35
boost/1_55_0(default)
boost/1_59_0
bowtie/0.12.7
bowtie/0.12.8(default)
bowtie2/2.1.0
bowtie2/2.2.5(default)
bwa/0.6.2
bwa/0.7.7(default)
castep/5.5(default)
cern_root/5.26.00
cern_root/5.30.00(default)
cern_root/5.32.01
cern_root/5.34.18
cernlib/2006(default)
cLhep/2.1.0.1(default)
cLhep/2.1.1.0
cLhep/2.1.3.1
cmake/2.8.7
cmake/3.0.2(default)
comsol/4.3a
comsol/4.3b
comsol/4.4
comsol/5.0
comsol/5.2(default)
cuda/4.0(default)
cuda/4.1
cuda/4.2
cuda/5.0
cuda/5.5
cuda/7.0
cufflinks/2.0.2
cufflinks/2.1.1(default)
cufflinks/2.2.1
ddd/3.3.12(default)
ddplot/2.5
ddplot/4.0(default)
eclipse/3.6
eclipse/4.3(default)
espresso/5.0.3
fastx_toolkit/0.0.13.2
ffmpeg/0.11.1
ffmpeg/1.1.1(default)
freeBayes/17Feb12
freetds/0.91
fribidi/0.19.6
gatk/1.0.4
gatk/2.7.2
gatk/3.1.1
gatk/3.3.0(default)
gcc/4.3.5
gcc/4.4(default)
gcc/4.7.2
gcc/4.9.3
gcta/0.93.9
gdal/1.9.2
gdc/0.30
gdc/4.9.3(default)
geant4/10.00.p02_wG4ML
geant4/4.9.3.p02(default)
geant4/4.9.4.p02
geant4/4.9.5
geant4/9.6.2
geant4/9.6.p02_wG4ML
geos/3.4.2(default)
gftp/2.0.19
ghc/7.6.3
gmp/4.3.2
gnuplot/4.2.3
gnuplot/4.4.4(default)
gotoblas2/1.13_multi_threaded
gotoblas2/1.13_single_threaded
gpac/0.5.1
grabs/2.0.1(default)
graphviz/2.28.0(default)
gromacs/4.6.5
handbrake/0.10.0(default)
handbrake/0.9.8
haskell/2013.2.0.0
hdfs/1.8.11_intel13.cs
hdfs/1.8.11_intel13.cs_intelmpi4.1.1
hdfs/1.8.14_intel-13.1.1_intelmpi-4.1.1
homer/4.7
hyperworks/11.0
hyperworks/12.0(default)
idl/6.3
idl/8.1(default)
idr/2010-10
impute/2.3.0(default)
intel/11.1
intel/12.0
intel/12.1
intel/13.0
intel/13.cs(default)
intel/14.cs
intelmpi/4.1.1
intelmpi/4.1.3
intelmpi/5.0.0
jags/3.3.0
jags/3.4.0(default)
java/1.6.0_23(default)
java/1.7.0_45
julia/0.3.11
lammps/10Feb15-intel14.0-imp15.0(default)
lammps/22Feb13-intel11.1-openmpi1.4.5
lammps/28Jun14-intel13.1-imp14.1
lynx/2.8.7(default)
maple/16
maple/17(default)
maq/0.7.1(default)
mathematica/10.0
mathematica/10.3(default)
mathematica/8.0.4
mathematica/9.0
matlab/7.11
matlab/7.14
matlab/7.7
matlab/8.2
matlab/8.4
matlab/8.6(default)
mats/3.0.7
mecab/0.996
mono/2.8(default)
mopac/2012-cpu
mpc/0.8.1
mpfr/2.4.2
mplayer
nag/5.3
ncl/5.2.1(default)
nco/4.0.6(default)
nco/4.4.4
netbeans/7.4
netbeans/8.0.2(default)
netcdf/4.1.3(default)
netcdf/4.1.3-shared
netcdf/4.1.3_i
netcdf/4.2.3-c
netcdf/4.4.2-fortran
ngsplot/2.47
nchen/6.5(default)
octave/3.6.1
octave/3.6.4(default)
openbabel/2.3.1(default)
openmpi/1.6_gcc-4.4
opensees/2.4.1
opensees/2.4.3
opensees/2.4.4(default)
opensees/2.4.4_parallel
paraview/3.6.2(default)
perl/5.10.1(default)
pgi_compiler/15.3
pgsl/0.1
plink/1.08
pop-c++/1.3(default)
preseq/1.0.2
proj/4.8.0
pypy/1.9
pypy3/2.4.0
python/2.6(default)
python/2.7
python/2.7.3
python/3.1
python/3.4
qcachegrind/0.7.4
qchem/3201s
qchem/3202mpich
qchem/3202s
qchem/4.3
qchem/4.3mpich(default)
qchem/4002mpich
qchem/4002s
qchem/4101
qchem/4101mpich
qiime/1.8.0
qiime/1.9.0(default)
relicon/1.3
ruby/1.9.2(default)
samtools/0.1.17
samtools/0.1.18
samtools/0.1.19(default)
samtools/1.2
seqlogo/2.8.2
shapelt/1
solar/4.3.1(default)
soxr/0.1.0
splicetrap/0.90.5
stata/11
stata/13
stata/14(default)
svn/1.6(default)
tcltk/8.4_thread-disabled(default)
tecplot360/2014
tecplot360/2015(default)
texlive/2012(default)
texlive/1.3(default)
tophat/1.3.3
tophat/2.0.14(default)
tophat/2.0.4
tophat/2.0.9
treemix/1.12
trinity/2013-08-14
vcftools/0.1.12a
vcftools/0.1.14(default)
vcftools/0.1.9
vegas/0.8.27
vmd/1.8.7(default)
vtk/5.8(default)
x264/000(default)
x264/20141209-2245
yasm/1.2.0(default)
zlib/1.2.8
```

Useful Tools on Hoffman2

ls /u/local/apps/

```
[flay@login3 ~]$  
[flay@login3 ~]$ ls /u/local/apps/  
abaqus          blcr             emacs            grads            libgtextutils   mumps            picard-tools     scilab           tvmet  
abaqusdocs      blitz            espresso         graphicsmagick  libtool          muscle           plink            scones           uclust  
accelrys        bmapuclatools   fastphase        graphviz         libxc            mygroup          pop-c++          sentaurus        udunits  
Accelrys        boost           fasttree         gromacs          lmbd             mysql            povray           shapeit          usearch  
activeperl      boost-jam        fastx_toolkit    gsl              loni_pipeline    namd             pp1cer           shapelib         usr_lib_GLOverride  
ActiveTcl       bowtie           fbat             hadoop           lumeral          ncl              preseq           skampi           valgrind  
adina           bowtie2          fe-safe          handbrake        lynx              nco              proj.4           slots            vasp  
Adobe           bwa             ffmpeg           harminv          mach              netbeans         protobuf         snappy           vcftools  
affymetrix      caffe           fftw2            haskell          mafft            netcdf           pypy            solar            vegas  
AFNI            casava           fftw3            hdf              manorm           newbler          pypy3           sortmerna        ViennaRNA  
aida            cdbtools        flex             hdf5             maple            ngsplot          python           soxr             vim  
amber           cd-hit           fltk             homer            maq              nlopt            qcachegrind      splicetraps      visit  
ampliconnoise   cernlib         freebayes        hyperworks       maqview          numpy            qchem            spm              vmd  
anaconda        cern_root       freesurfer       iaida            mathematica       nwchem           qgfe             stata            votca  
annovar         clearcut        freetds          idl              matio            ocaml            qhull            stressapptest    vtk  
ansys_inc       clhep           frbidi           idr              matlab           octave           qiime             structure         wannier90  
ant             cln             fsl              igraph           mats             omssa            qiime_data        subversion        weblogo  
antlr           cmake           gamess           ilog             mecab            oommf            qrupdate          suitesparse       wolfram  
armadillo       common          gatk             IM-IMA           meep             openbabel         qsub             sundials          x264  
arpack          consol          gatk-queue       impute           meld             openfoam          superlu_dist      x86_open64  
atlas           conan           gaussian          imsl             merlin           openmotif         swarm             xerces-c  
atomeye         condor          gaussview         infernal          metis            openpensee        sysstat           xfdtd  
atompaw         consed          gcta             inspect          mfold            osiris            R                 xfig_OLD  
autoconf        cp2k            gdal             installation     microbioemutil   osu-micro-benchmarks  raxml            xilinx-vivado  
automake        cpmd            gdc              intel            migrate           papi              rdp_classifier    tau              xmd  
bamtools        ctffind         geant4           insight          minirosetta       paraview          relion            tc1              xmedcon  
bcftools        cufflinks       genotorrent      jags             mira              parmetis          remcom            tcsh             xpdf_OLD  
bcl2qfastq      cytoscape       geos             jam              molcas            parsec            RepeatModeler     texlive          yam1  
beagle          ddd             gflags           java              molder            parsinsert         resmap            tmux             yasm  
beaglecall      ddplot          gftp             jaxodraw         mono              pcre              rosetta           tophat           zlib  
beagle_utilities dejagnu          git_OLD          jmol             mopac              pdt               rstudio           totalview  
bedtools        dirac           globalarrays     kepler           mothur             perl_modules      rsync             trans-ABYSS      treemix  
bfast           diskusage       glog             lammps           mpf                petsc             ruby              trilos           trinity  
bioscope        dx              gmp              lapack            mpblast            phantompeakqualtools  sage              tt  
blas            eclipse          gpac             lapack++          mplayer  
blast+          eigen           grace            ldope  
blat            eigensoft       leveldb
```


Useful Tools on Hoffman2 Galaxy

```
cd /u/home/galaxy/collaboratory/apps/
```

```
[flay@login4 apps]$ ls -lh
```

```
total 512K
```

drwxr-sr-x	5	flay	collabor	4.0K	Apr	6	16:37	fit-hi-c
drwxr-sr-x	3	sorel	collabor	4.0K	Apr	18	14:47	gatk
drwxr-sr-x	6	flay	collabor	4.0K	Apr	6	11:48	HiCPlotter
drwxr-sr-x	3	flay	collabor	4.0K	Apr	6	09:40	hic-pro
drwxr-sr-x	7	flay	collabor	4.0K	Apr	6	09:40	HiC-Pro
drwxr-sr-x	6	flay	collabor	4.0K	Oct	23	2015	hicup_v0.5.8
drwxr-xr-x	7	flay	collabor	16K	Apr	5	20:04	samtools-1.1

Adding Tools to PATH

#what is in your environment?

echo \$PATH

```
[flay@login4 hic-pro]$ echo $PATH
/u/local/compilers/intel-cs/2013.0.028/bin:/u/local/compilers/intel-cs/2013.0.028/advisor_xe/bin64:/u/local/compilers/intel-cs/2013.0.028/vtune_amplifier_xe/bin64:/u/local/compilers/intel-cs/2013.0.028/inspector_xe/bin64:/u/local/compilers/intel-cs/2013.0.028/mpi/intel64/bin:/u/local/compilers/intel-cs/2013.0.028/itac/intel64/bin:/u/local/compilers/intel-cs/2013.0.028/compiler_xe/bin/intel64:/u/local/compilers/intel-cs/2013.0.028/composer_xe/mpirt/bin/intel64:/u/local/compilers/intel-cs/2013.0.028/composer_xe/bin/intel64_mic:/u/local/compilers/intel-cs/2013.0.028/composer_xe/debugger/gui/intel64:/u/systems/UGE8.0.1vm/bin/lx-amd64:/u/local/bin:/u/local/sbin:/usr/lib64/qt-3.3/bin:/usr/local/bin:/bin:/usr/bin:/usr/local/sbin:/usr/sbin:/sbin:/u/home/f/flay/bin:/u/home/f/flay/software/biscuit/bin:/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin:/u/home/galaxy/collaboratory/apps/hicup_v0.5.8:/u/home/f/flay/software/metilene_v0.2-5:/u/home/f/flay/software/mcl-14-137/bin:/u/home/f/flay/software/bwa-meth:/u/home/f/flay/software/Bis-tools/External_tools/ucsc_tools:/u/home/f/flay/software/bsmap-2.90
```

#what shell are you using?

echo \$SHELL

```
[flay@login4 hic-pro]$ echo $SHELL
/bin/bash
```

Adding Tools to PATH

##HiC-Pro

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin

```
[flay@login4 bin]$ ls  
HiC-Pro  utils
```

##hicup

/u/home/galaxy/collaboratory/apps/hicup_v0.5.8

```
[flay@login4 hicup_v0.5.8]$ ls  
config_files  hicup  hicup_filter  hicup_report.html  README.txt  
Conversion    hicup_deduplicator  hicup_mapper  hicup_truncater  RELEASE_NOTES.txt  
Documentation  hicup_digester      hicup_module.pm  LICENSE.txt      r_scripts
```

Adding Tools to PATH

#add the directories of tools in to your environment for quicker access

```
#the following example is for if you are in bash
vi ~/.bash_profile
```

```
# .bash_profile

# Get the aliases and functions
if [ -f ~/.bashrc ]; then
    . ~/.bashrc
fi

# User specific environment and startup programs


PATH=$PATH:$HOME/bin:/u/home/f/flay/software/biscuit/bin:/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin:/u/home/galaxy/collaboratory/apps/hicup_v0.5.8:/u/home/f/flay/software/melenc-v0.2-5:/u/home/f/flay/software/mcl-14-137/bin:/u/home/f/flay/software/bwa-meth:/u/home/f/flay/software/Bis-tools/External_tools/ucsc_tools:/u/home/f/flay/software/bsmap-2.90

export PATH
export LD_LIBRARY_PATH=$HOME/.local/lib/:$LD_LIBRARY_PATH
export BISTOOLS=/u/home/f/flay/software/Bis-tools
export HDF5_LIBRARIES=/u/local/apps/hdf5/1.8.14/intel-13.1.1/intelmpi-4.1.1/lib

~
~
~
~
~
~
~
~
~
~
~
```

#Add the path to HiC-Pro and hicup

Type i to insert
Once added, press esc

##to save, type
:wq!

"~/ .bash_profile" 15L, 714C

15,1 All

```
#Add the path to HiC-Pro and  
/intelmpi-4.1.1/lib  
hicup
```

Type i to insert
Once added, press esc

```
##to save, type
:wq!
```

Adding Tools to PATH

#Two methods to save changes to your environment:

1. After exiting `.bash_profile`, type:

`source ~/.bash_profile`

2. Log out of Hoffman2 session and log back in. The changes should be saved.

3. To check, type:

`echo $PATH`

or call, HiC-Pro directly:

`HiC-Pro`

```
[flay@login4 flay]$ HiC-Pro
usage : HiC-Pro -i INPUT -o OUTPUT -c CONFIG [-s ANALYSIS_STEP] [-p] [-h] [-v]
Use option -h|--help_ for more information
```

Alignment Using HiC-Pro: On Hoffman2 Interactively

#grab a Hoffman2 session:

qrsh

or

qrsh -now n -l i,mem=5G,time=02:00:00,exclusive=TRUE-pe shared 6

Request exact configuration

time limit of the session
(default is 2 hours,
24hrs max)

Reserve the whole node

Request 6 cores

Memory requested per
core (default is 1G)

Alignment Using HiC-Pro: Copy Files

#copy all the files you need to scratch folder

cd /u/scratch/f/flay/ change to your ID

cp /u/scratch/f/flay/workshop10.tar.gz . Indicates
current
directory

tar -xvzf workshop10.tar.gz

cd workshop10

ls

Alignment Using HiC-Pro: Loading Modules

##Load all the necessary modules:

##must use the exact versions

module load python/2.7.3

module load R/3.2.3

module load bowtie2

module load samtools/1.2

##check if you have 'iced' python package, by typing:

python

import iced

##if iced is not already installed, install by typing the following on command line:

pip install iced --user

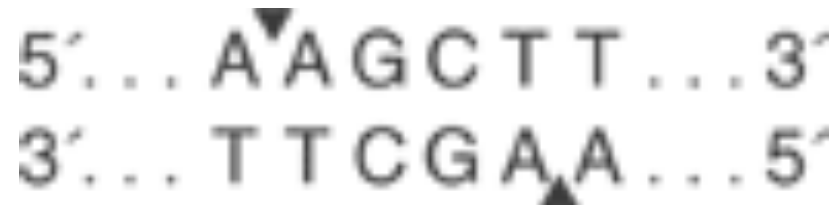
Alignment Using HiC-Pro: Prepare Digested Fragment File

##Find the recognition sites of restriction enzyme

##This example dataset uses HindIII

<https://www.neb.com/products/restriction-endonucleases>

HindIII



Alignment Using HiC-Pro: Prepare Digested Fragment File

##Use digest_genome.py utility in HiC-Pro to create a .bed file containing all digested fragments

```
[flay@login1 ~]$ python /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin/utils/digest_genome.py -h
usage: digest_genome.py [-h] [-r RES_SITES [RES_SITES ...]] [-o OUT] fastafile
```

positional arguments:
fastafile

optional arguments:

```
-h, --help            show this help message and exit
-r RES_SITES [RES_SITES ...], --restriction_sites RES_SITES [RES_SITES ...]
                        The cutting position has to be specified using '^'.
                        For instance, -r A^GATCT for HindIII digestion.
                        Several restriction enzyme can be specified.

-o OUT, --out OUT
```

Alignment Using HiC-Pro: Prepare Digested Fragment File

```
python /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin/utis/digest_genome.py -r A^AGCTT -o hic-hg19_HindIII.bed /u/scratch/f/flay/workshop10/genome/hg19_rCRSchrM.fa
```



A terminal window titled "flay@n2180: /u/scratch/f/flay/workshop10 — ssh — 125x33" displays the execution of the HiC-Pro script. The user enters the command to digest the hg19 genome with HindIII. The script outputs the restriction site (AAGCTT) and the offset (1). It then lists 15 chromosomes (chr1 to chr15) and the number of random fragments generated for each: chr1 (191), chr1 (192), chr2, chr3, chr4, chr4 (193), chr4 (194), chr5, chr6, chr7, chr7 (195), chr8, chr8 (196), chr8 (197), chr9, chr9 (198), chr9 (199), chr9 (200), chr9 (201), chr10, chr11, chr11 (202), chr12, chr13, chr14, and chr15.

```
flay@n2180 workshop10]$  
[flay@n2180 workshop10]$ python /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/bin/utis/digest_genome.py -r A^AGCTT  
-o hic-hg19_HindIII.bed /u/scratch/f/flay/workshop10/genome/hg19_rCRSchrM.fa  
Analyzing /u/scratch/f/flay/workshop10/genome/hg19_rCRSchrM.fa  
Restriction site(s) AAGCTT  
Offset(s) 1  
chr1 ...  
chr1_gl000191_random ...  
chr1_gl000192_random ...  
chr2 ...  
chr3 ...  
chr4 ...  
chr4_gl000193_random ...  
chr4_gl000194_random ...  
chr5 ...  
chr6 ...  
chr7 ...  
chr7_gl000195_random ...  
chr8 ...  
chr8_gl000196_random ...  
chr8_gl000197_random ...  
chr9 ...  
chr9_gl000198_random ...  
chr9_gl000199_random ...  
chr9_gl000200_random ...  
chr9_gl000201_random ...  
chr10 ...  
chr11 ...  
chr11_gl000202_random ...  
chr12 ...  
chr13 ...  
chr14 ...  
chr15 ...
```

Alignment Using HiC-Pro: Prepare Digested Fragment File

how many HindIII fragments are generated?

```
[flay@n2180 workshop10]$ wc -l hic-hg19_HindIII.bed
839263 hic-hg19_HindIII.bed
```

what do the output file look like?

```
[flay@n2180 workshop10]$ head hic-hg19_HindIII.bed
chr1      0      16007  HIC_chr1_1      0      +
chr1     16007     24571  HIC_chr1_2      0      +
chr1     24571     27981  HIC_chr1_3      0      +
chr1     27981     30429  HIC_chr1_4      0      +
chr1     30429     32153  HIC_chr1_5      0      +
chr1     32153     32774  HIC_chr1_6      0      +
chr1     32774     37752  HIC_chr1_7      0      +
chr1     37752     38369  HIC_chr1_8      0      +
chr1     38369     38791  HIC_chr1_9      0      +
chr1     38791     39255  HIC_chr1_10     0      +
```

Alignment Using HiC-Pro: Configure Alignment Settings

config-hicpro.txt file must be specified

##to edit, type on command line:

view config-hicpro.txt

##alternatively, download the file on to your desktop, open and edit using your text editor of choice.

##detailed description for each parameter is discussed here:

<http://nservant.github.io/HiC-Pro/MANUAL.html>

Alignment Using HiC-Pro: Configure Alignment Settings

flay@login1:/u/scratch/f/flay/workshop10 — ssh — 156x41

```
#####  
## Paths and Settings - Do not edit !  
#####
```

```
TMP_DIR = tmp  
LOGS_DIR = logs  
BOWTIE2_OUTPUT_DIR = bowtie_results  
MAPC_OUTPUT = hic_results  
RAW_DIR = rawdata
```

```
#####  
## SYSTEM AND SCHEDULER - Start Editing Here !!  
#####  
N_CPU = 2  
LOGFILE = hicpro.log
```

```
JOB_NAME =  
JOB_MEM =  
JOB_WALLTIME =  
JOB_QUEUE =  
JOB_MAIL =
```

```
#####  
## Data  
#####
```

```
PAIR1_EXT = _R1  
PAIR2_EXT = _R2
```

```
#####  
## Alignment options  
#####
```

```
FORMAT = phred33  
MIN_MAPQ = 0
```

```
BOWTIE2_IDX_PATH = /u/scratch/f/flay/workshop10/genome/  
BOWTIE2_GLOBAL_OPTIONS = --very-sensitive -l 30 --score-min L,-0.6,-0.2 --end-to-end --reorder  
BOWTIE2_LOCAL_OPTIONS = --very-sensitive -L 20 --score-min L,-0.6,-0.2 --end-to-end --reorder  
[]
```

#follow naming convention of
your raw fastq files

#change to the directory
where your bwt2.index is
stored

Alignment Using HiC-Pro: Configure Alignment Settings

flay@login1:/u/scratch/f/flay/workshop10 — ssh — 156x41

```
#####  
## Annotation files  
#####  
  
REFERENCE_GENOME = hg19  
GENOME_SIZE = /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes  
  
#####  
## Allele specific analysis  
#####  
  
ALLELE_SPECIFIC_SNP =  
  
#####  
## Digestion Hi-C  
#####  
  
GENOME_FRAGMENT = /u/scratch/f/flay/workshop10/hic-hg19_HindIII.bed  
LIGATION_SITE = AAGCTAGCTT  
MIN_FRAG_SIZE = 100  
MAX_FRAG_SIZE = 100000  
MIN_INSERT_SIZE = 100  
MAX_INSERT_SIZE = 800  
  
#####  
## Hi-C processing  
#####  
  
MIN_CIS_DIST =  
GET_ALL_INTERACTION_CLASSES = 1  
GET_PROCESS_SAM = 0  
RM_SINGLETON = 1  
RM_MULTI = 1  
RM_DUP = 1  
  
#####  
## Contact Maps  
#####
```

#change to the directory where
your chrom sizes file is stored

#specify the location and name of
digested fragment file
#specify ligation junction
#specify fragment and insert sizes

Alignment Using HiC-Pro: Configure Alignment Settings

```
#####  
## Contact Maps  
#####  
BIN_SIZE = 20000 40000 150000 500000 1000000  
MATRIX_FORMAT = upper  
#####  
## Normalization  
#####  
MAX_ITER = 100  
FILTER_LOW_COUNT_PERC = 0.02  
FILTER_HIGH_COUNT_PERC = 0  
EPS = 0.1
```

#bin size of contact matrix

##save edited config file

##make sure they are in your working directory before
proceeding

Alignment Using HiC-Pro: Run HiC-Pro

##to see the HiC-Pro command options, type:
HiC-Pro -h

```
[flay@login1 ~]$ HiC-Pro -h
usage : HiC-Pro -i INPUT -o OUTPUT -c CONFIG [-s ANALYSIS_STEP] [-p] [-h] [-v]
Use option -h|--help for more information
```

HiC-Pro 2.7.6

OPTIONS

```
-i|--input INPUT : input data folder; Must contains a folder per sample with input files
-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 specified the complete workflow is run
    mapping: perform reads alignment
    proc_hic: perform Hi-C filtering
    quality_checks: run Hi-C quality control plots
    merge_persample: merge multiple inputs and remove duplicates if specified
    build_contact_maps: run merge_persample and build raw inter/intrachromosomal contact maps
    ice_norm : run ICE normalization on contact maps
[-h|--help]: help
[-v|--version]: version
```

Alignment Using HiC-Pro: Run HiC-Pro

HiC-Pro -i /u/scratch/f/flay/workshop10/data -o output -c config-hicpro.txt

```
[flay@n7199 workshop10]$ HiC-Pro -i /u/scratch/f/flay/workshop10/data -o output -c config-hicpro.txt
```

Run HiC-Pro 2.7.6

Sat May 7 18:13:40 PDT 2016

Bowtie2 alignment step1 ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/bowtie_wrap.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt -u >> hicpro.log

Sat May 7 18:14:42 PDT 2016

Bowtie2 alignment step2 ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/bowtie_wrap.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt -l >> hicpro.log

Sat May 7 18:14:48 PDT 2016

Combine both alignment ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/bowtie_combine.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:14:49 PDT 2016

Bowtie2 mapping statistics for R1 and R2 tags ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/mapping_stat.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:14:49 PDT 2016

Pairing of R1 and R2 tags ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/bowtie_pairing.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:14:54 PDT 2016

Assign alignments to restriction fragments ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/mapped_2hic_fragments.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:15:08 PDT 2016

Merge multiple files from the same sample ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/merge_valid_interactions.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:15:08 PDT 2016

Merge stat files per sample ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/merge_stats.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log

Sat May 7 18:15:09 PDT 2016

Run quality checks for all samples ...

/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/make_plots.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt -p "all" >> hicpro.log

Alignment Using HiC-Pro: Run HiC-Pro

```
Sat May 7 18:15:14 PDT 2016
Generate binned matrix files ...
/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_raw_maps.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt
cat hic_results/data//rawdata/rawdata_allValidPairs | /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_matrix --binsize 20000 --chrsize
s /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes --ifile /dev/stdin --oprefix hic_results/matrix/rawdata/raw/20000/rawdata_20000 --matrix-format upper
2> logs/rawdata/build_raw_maps.log
cat hic_results/data//rawdata/rawdata_allValidPairs | /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_matrix --binsize 40000 --chrsize
s /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes --ifile /dev/stdin --oprefix hic_results/matrix/rawdata/raw/40000/rawdata_40000 --matrix-format upper
2> logs/rawdata/build_raw_maps.log
cat hic_results/data//rawdata/rawdata_allValidPairs | /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_matrix --binsize 150000 --chrsize
s /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes --ifile /dev/stdin --oprefix hic_results/matrix/rawdata/raw/150000/rawdata_150000 --matrix-format upper
2> logs/rawdata/build_raw_maps.log
cat hic_results/data//rawdata/rawdata_allValidPairs | /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_matrix --binsize 500000 --chrsize
s /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes --ifile /dev/stdin --oprefix hic_results/matrix/rawdata/raw/500000/rawdata_500000 --matrix-format upper
2> logs/rawdata/build_raw_maps.log
cat hic_results/data//rawdata/rawdata_allValidPairs | /u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/build_matrix --binsize 1000000 --chrsize
s /u/scratch/f/flay/workshop10/genome/hg19.chrom.sizes --ifile /dev/stdin --oprefix hic_results/matrix/rawdata/raw/1000000/rawdata_1000000 --matrix-format upper
2> logs/rawdata/build_raw_maps.log
```

```
Sat May 7 18:15:15 PDT 2016
Run ICE Normalization ...
/u/home/galaxy/collaboratory/apps/hic-pro/HiC-Pro_2.7.6/scripts/ice_norm.sh -c /u/scratch/f/flay/workshop10/config-hicpro.txt >> hicpro.log
[flay@n7199 workshop10]$
```