

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 2

Alignment Using HiC-Pro: HiC-Pro Output

##output directory is created

```
[flay@n2004 workshop10]$ ls
```

```
config-hicpro.txt  genome  hic-hg19_HindIII.bed  output  rawdata
```

##enter the output directory to see subdirectories

```
[flay@n2004 workshop10]$ cd output/
```

```
[flay@n2004 output]$ ls
```

```
bowtie_results  config-hicpro.txt  hicpro.log  hic_results  logs  rawdata  tmp
```

Alignment Using HiC-Pro: HiC-Pro Output

##alignment file is contained within bowtie_results/

```
[flay@n2004 output]$ cd bowtie_results/  
[flay@n2004 bowtie_results]$ ls  
bwt2 bwt2_global bwt2_local
```

##enter directory

##what files are created?

```
[flay@n2004 bowtie_results]$ cd bwt2/rawdata/  
[flay@n2004 rawdata]$ ls -lh  
total 14M  
-rw-r--r-- 1 flay matteop 536 May 2 22:19 rawdata.mpairstat  
-rw-r--r-- 1 flay matteop 265 May 2 22:19 rawdata__R1.mmapstat  
-rw-r--r-- 1 flay matteop 265 May 2 22:19 rawdata__R2.mmapstat  
-rw-r--r-- 1 flay matteop 4.5M May 2 22:19 test_dataset_hg19.bwt2pairs.bam  
-rw-r--r-- 1 flay matteop 315 May 2 22:19 test_dataset_hg19.bwt2pairs.pairstat  
-rw-r--r-- 1 flay matteop 3.6M May 2 22:19 test_dataset_R1_hg19.bwt2merged.bam  
-rw-r--r-- 1 flay matteop 89 May 2 22:19 test_dataset_R1_hg19.mapstat  
-rw-r--r-- 1 flay matteop 3.6M May 2 22:19 test_dataset_R2_hg19.bwt2merged.bam  
-rw-r--r-- 1 flay matteop 89 May 2 22:19 test_dataset_R2_hg19.mapstat
```

Alignment Using HiC-Pro: HiC-Pro Output

change directory to see hic_results

```
[flay@n2004 output]$ cd hic_results/  
[flay@n2004 hic_results]$ ls  
data  matrix  pic
```

data/ contains information on read pairs

```
[flay@n2004 hic_results]$ cd data/rawdata/  
[flay@n2004 rawdata]$ ls -lh  
total 11M  
-rw-r--r-- 1 flay matteop 2.9M May  2 22:19 rawdata_allValidPairs  
-rw-r--r-- 1 flay matteop 140 May  2 22:19 rawdata_allValidPairs.mergestat  
-rw-r--r-- 1 flay matteop 454 May  2 22:19 rawdata.mRSstat  
-rw-r--r-- 1 flay matteop 1.6M May  2 22:19 test_dataset_hg19.bwt2pairs.DEPairs  
-rw-r--r-- 1 flay matteop 1005K May  2 22:19 test_dataset_hg19.bwt2pairs.DumpPairs  
-rw-r--r-- 1 flay matteop 261 May  2 22:19 test_dataset_hg19.bwt2pairs.RSstat  
-rw-r--r-- 1 flay matteop 35K May  2 22:19 test_dataset_hg19.bwt2pairs.SCPairs  
-rw-r--r-- 1 flay matteop  0 May  2 22:19 test_dataset_hg19.bwt2pairs.SinglePairs  
-rw-r--r-- 1 flay matteop 2.9M May  2 22:19 test_dataset_hg19.bwt2pairs.validPairs
```

Alignment Using HiC-Pro: HiC-Pro Output

valid pairs data in tab-delimited text format

read name / chr_reads1 / pos_reads1 / strand_reads1 /
chr_reads2 / pos_reads2 / strand_reads2 / fragment_size [/
allele_specific_tag]

```
[flay@n2004 rawdata]$ head test_dataset_hg19.bwt2pairs.validPairs
SRR071233.15582277 chr1 774056 - chr18 58783466 + 251 HIC_chr1_200 HIC_chr18_17046 36 42
SRR071233.13850439 chr1 830494 + chr3 72610842 - 323 HIC_chr1_215 HIC_chr3_21488 36 42
SRR071233.13826897 chr1 848396 - chr11 48713429 + 324 HIC_chr1_219 HIC_chr11_14431 42 42
SRR071233.13192685 chr1 875945 + chr2 35386331 + 220 HIC_chr1_221 HIC_chr2_10032 42 42
SRR071233.13077187 chr1 943823 - chr2 239890283 + 209 HIC_chr1_228 HIC_chr2_70834 42 42
SRR071233.2710333 chr1 1069005 + chr10 13094125 + 347 HIC_chr1_241 HIC_chr10_3657 23 42
SRR071233.10914325 chr1 1151075 - chr1 8170636 - 373 HIC_chr1_251 HIC_chr1_1557 42 42
SRR071233.15210345 chr1 1307895 - chr20 37386273 + 106 HIC_chr1_261 HIC_chr20_9248 42 42
SRR071233.12538565 chr1 1426941 - chr13 42530754 - 235 HIC_chr1_277 HIC_chr13_7145 38 42
SRR071233.14589247 chr1 1465227 + chr5 5927515 - 334 HIC_chr1_282 HIC_chr5_1520 42 42
```

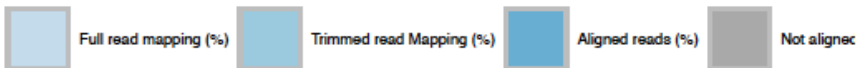
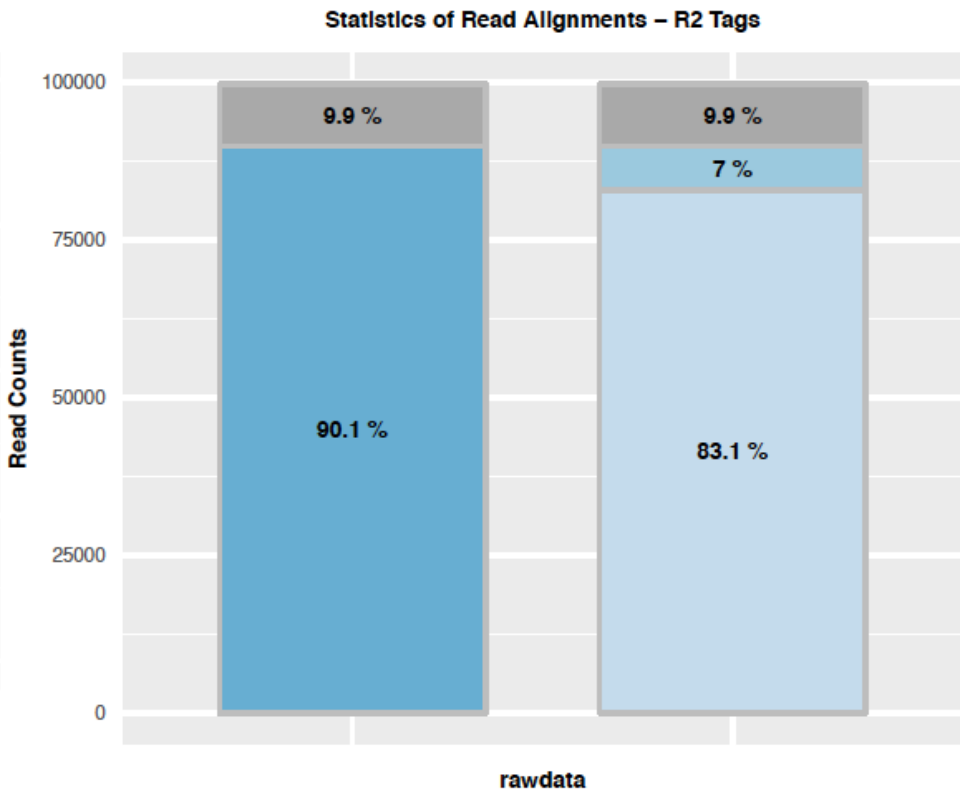
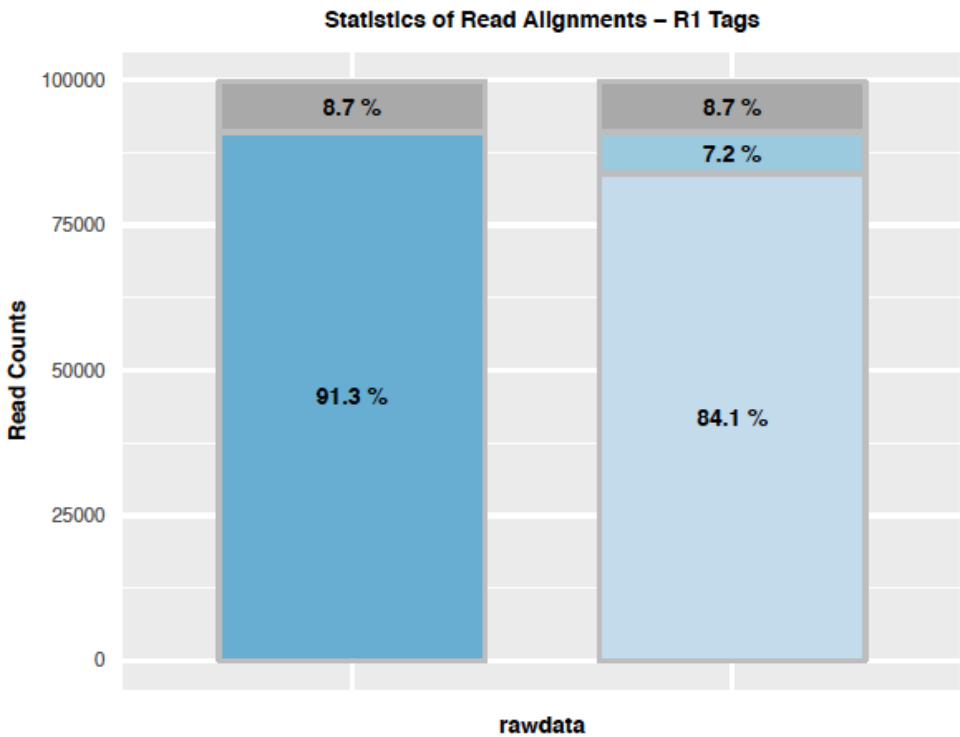
Alignment Using HiC-Pro: HiC-Pro Output

alignment summary statistics are contained within pic/

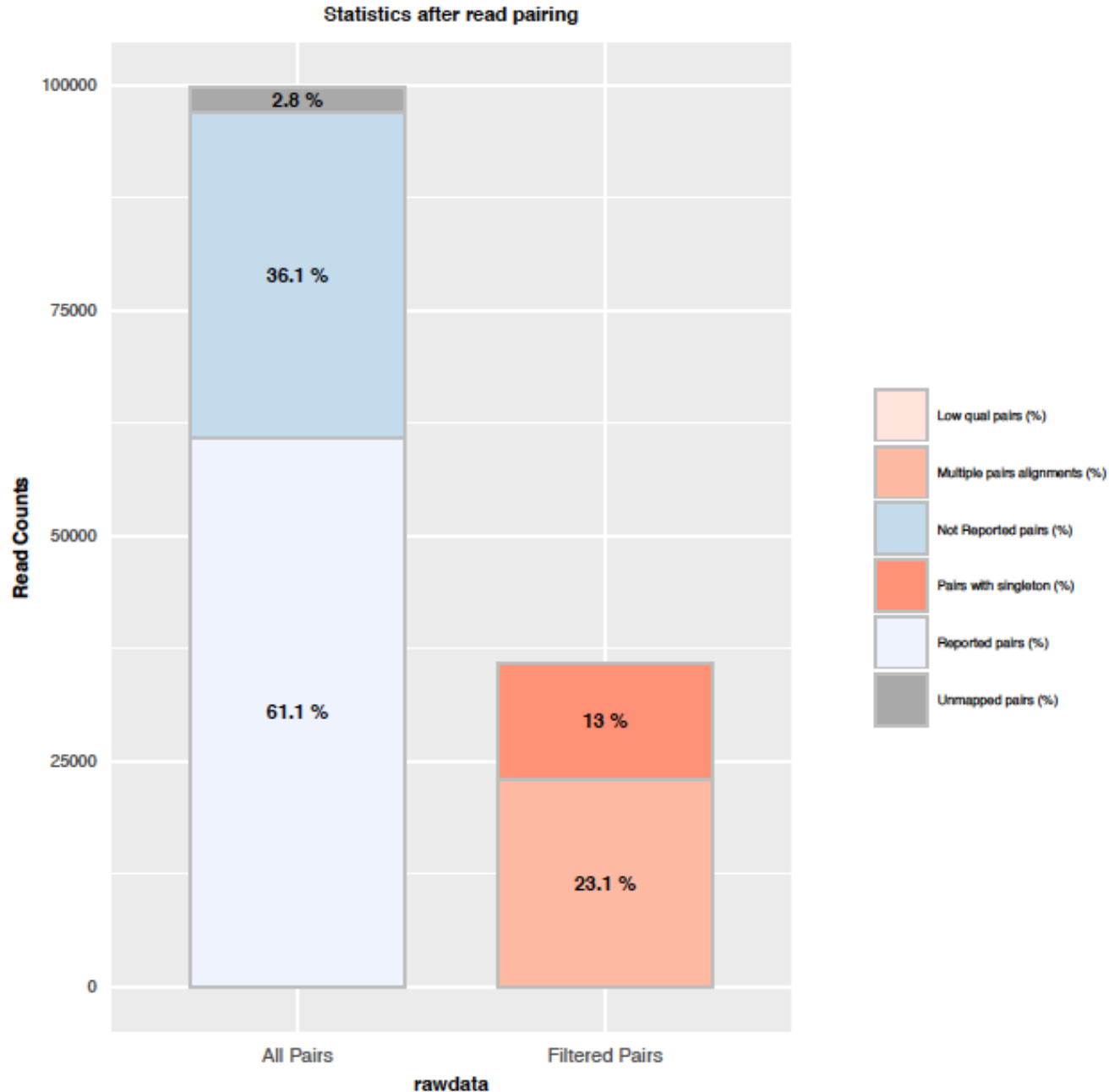
```
[flay@n2004 rawdata]$ ls  
plotHiCContactRanges_rawdata.pdf  plotHiCFragmentSize_rawdata.pdf  plotMapping_rawdata.pdf  
plotHiCFragment_rawdata.pdf       plotMappingPairing_rawdata.pdf
```

What does a succesful HiC experiment look like?

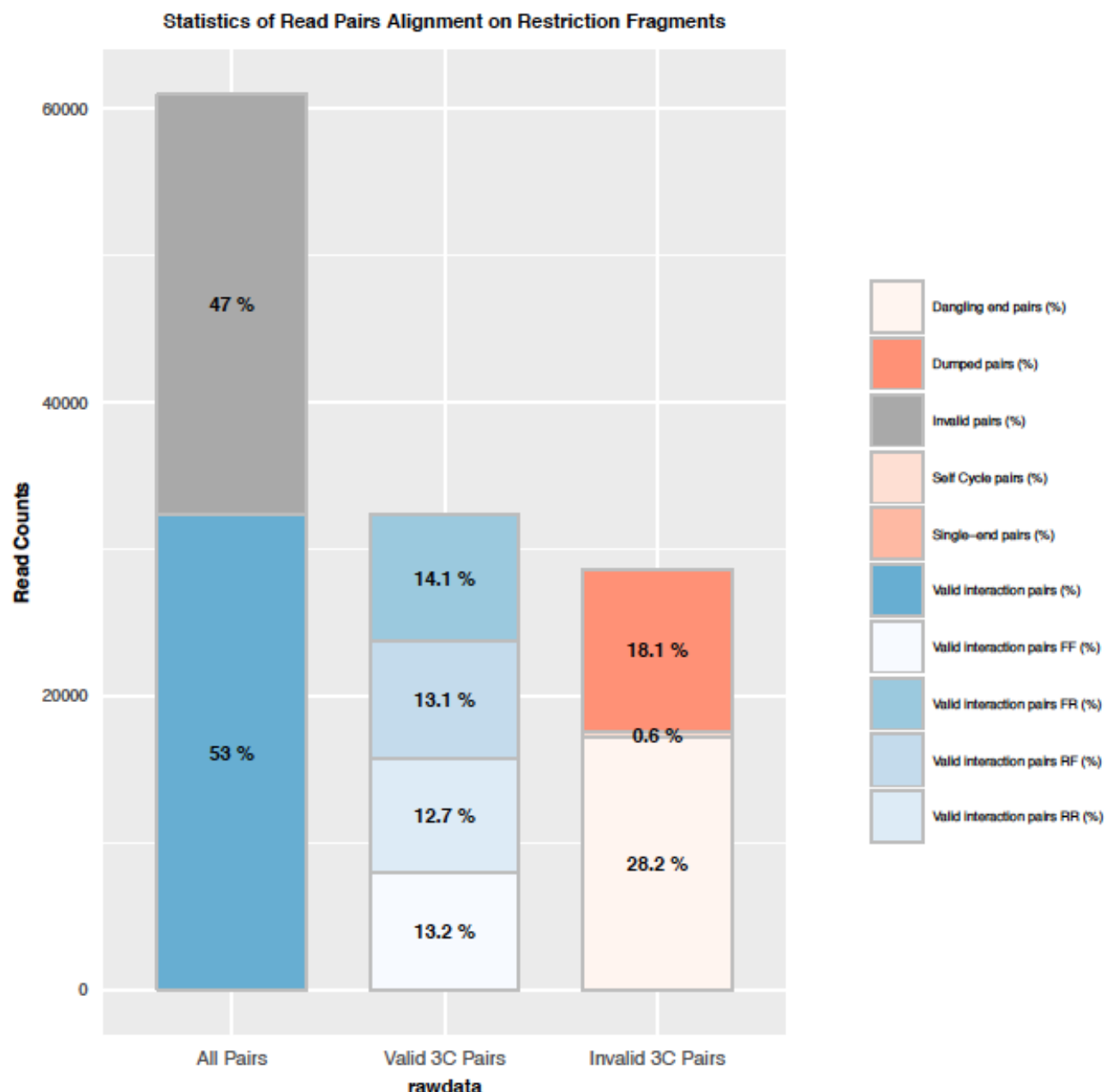
HiC-Pro Output: Mapping Statistics



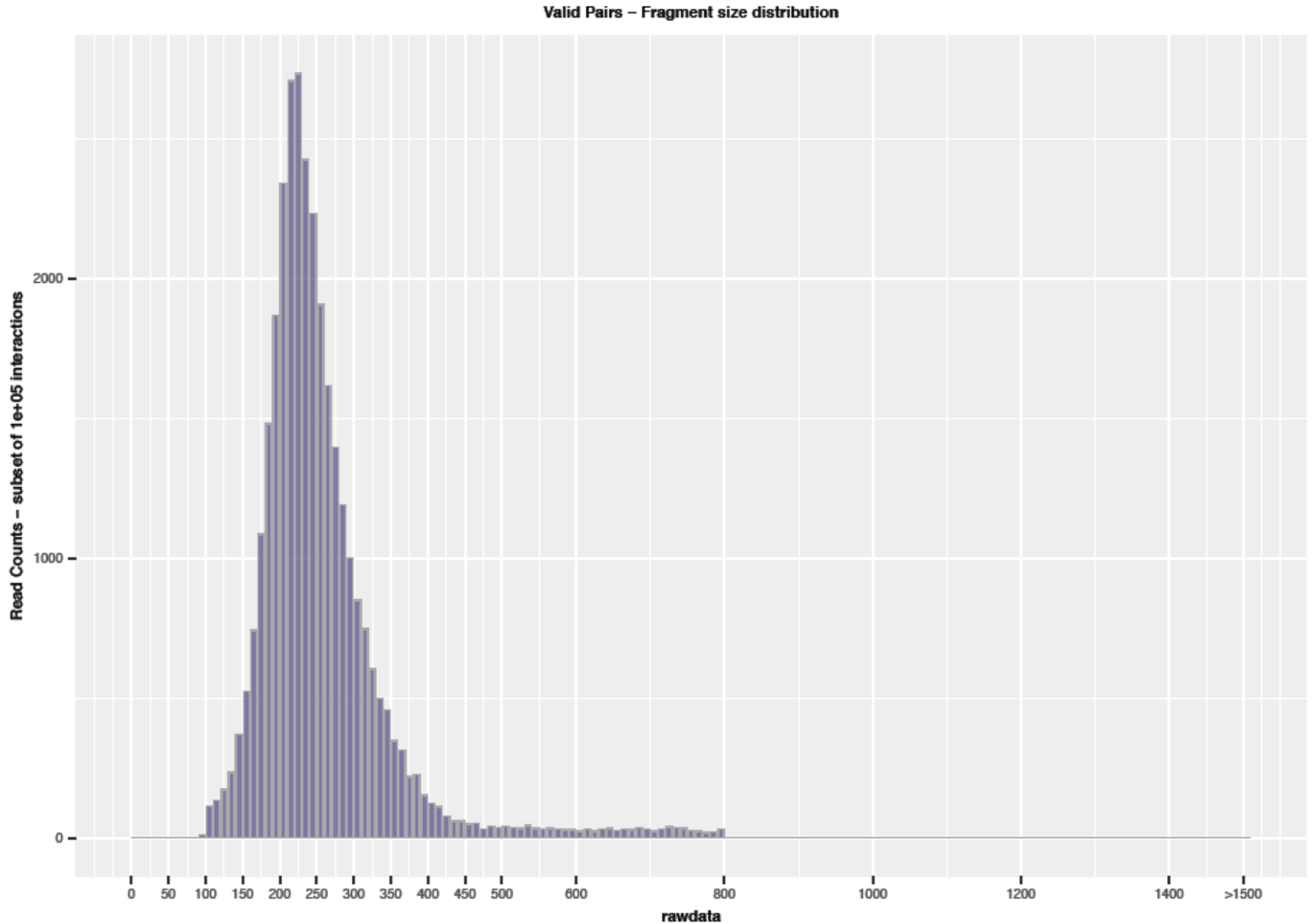
HiC-Pro Output: Raw Pairing Statistics



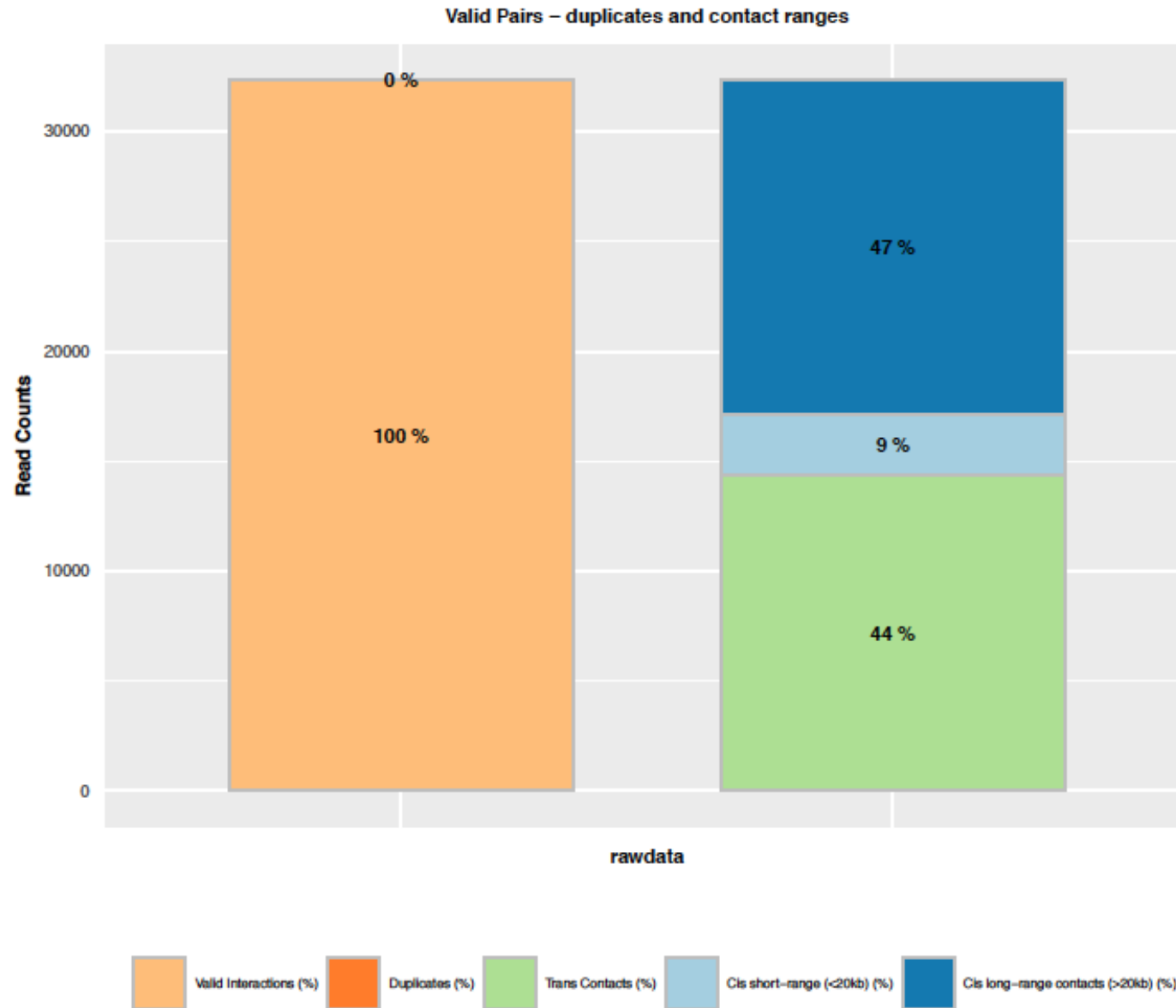
HiC-Pro Output: Read Pairing Statistics



HiC-Pro Output: Size Distribution



HiC-Pro Output: Contact Ranges



HiC-Pro Output: Contact Matrix

```
## in hic_results/matrix/rawdata/  
##normalized and raw contact matrices
```

```
[flay@n2004 hic_results]$ cd matrix/rawdata/  
[flay@n2004 rawdata]$ ls  
iced raw
```

```
## bin size as indicated in config-hicpro
```

```
[flay@n2004 rawdata]$ cd raw  
[flay@n2004 raw]$ ls  
1000000 150000 20000 40000 500000  
[flay@n2004 raw]$  
[flay@n2004 raw]$ cd 20000  
[flay@n2004 20000]$ ls  
rawdata_20000_abs.bed rawdata_20000.matrix rawdata_20000_ord.bed
```

HiC-Pro Output: Contact Matrix

output in triplet sparse format

```
[flay@n2004 20000]$ head rawdata_20000.matrix
```

39	139779	1
42	28254	1
43	100999	1
44	14233	1
48	24458	1
54	92441	1
58	409	1
66	142613	1
72	114134	1
74	44380	1

```
[flay@n2004 20000]$ head rawdata_20000_abs.bed
```

chr1	0	20000	1
chr1	20000	40000	2
chr1	40000	60000	3
chr1	60000	80000	4
chr1	80000	100000	5
chr1	100000	120000	6
chr1	120000	140000	7
chr1	140000	160000	8
chr1	160000	180000	9
chr1	180000	200000	10

Visualization of Contact Matrix

Akdemir and Chin *Genome Biology* (2015) 16:198
DOI 10.1186/s13059-015-0767-1



SOFTWARE

Open Access

HiCPlotter integrates genomic data with interaction matrices



Kadir Caner Akdemir¹ and Lynda Chin^{1,2*}

Abstract

Metazoan genomic material is folded into stable non-randomly arranged chromosomal structures that are tightly associated with transcriptional regulation and DNA replication. Various factors including regulators of pluripotency, long non-coding RNAs, or the presence of architectural proteins have been implicated in regulation and assembly of the chromatin architecture. Therefore, comprehensive visualization of this multi-faceted structure is important to unravel the connections between nuclear architecture and transcriptional regulation. Here, we present an easy-to-use open-source visualization tool, HiCPlotter, to facilitate juxtaposition of Hi-C matrices with diverse genomic assay outputs, as well as to compare interaction matrices between various conditions. <https://github.com/kcakdemir/HiCPlotter>

Visualization of Contact Matrix: Plot an Interaction Matrix of Chr8

```
##change directory to matrix/  
cd matrix/
```

##run the following command to plot raw matrix:

```
python /u/home/galaxy/collaboratory/apps/HiCPlotter/  
HiCPlotter.py -f rawdata_500000.matrix -bed  
rawdata_500000_abs.bed -n raw -chr chr8 -o raw_chr8 -tri 1  
-r 500000 -hmc 5 -mm 10 -ptr 1
```

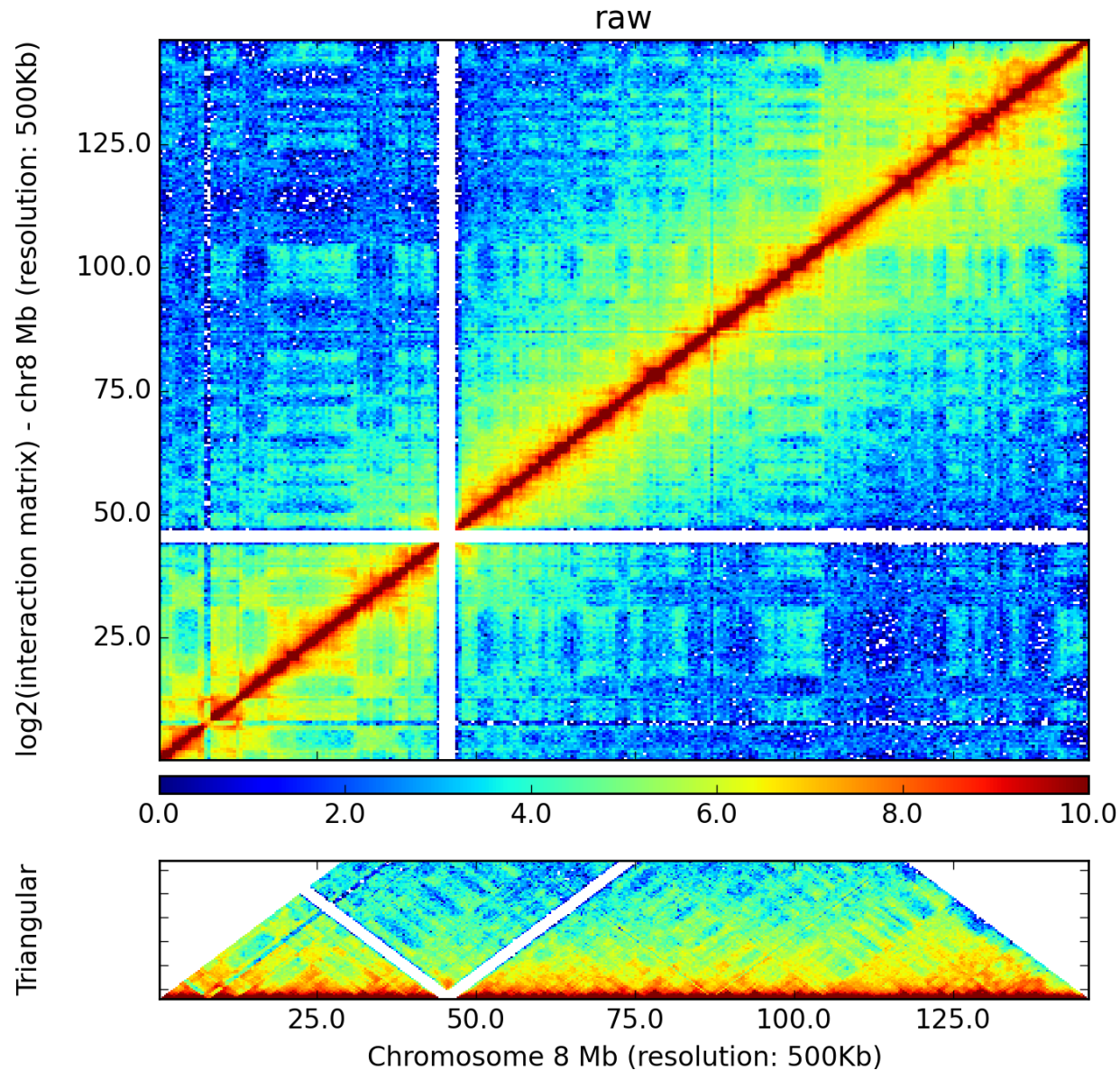
```
[flay@n2240 matrix]$ python /u/home/galaxy/collaboratory/apps/HiCPlotter/HiCPlotter.py -f rawdata_500000.matrix -bed rawdata_500000_abs.bed  
-n raw -chr chr8 -o raw_chr8 -tri 1 -r 500000 -hmc 5 -mm 10 -ptr 1  
Plotting now!!
```

##output file is in .png format

##download to your desktop and open

```
scp -r flay@hoffman2.idre.ucla.edu:/u/scratch/f/flay/  
workshop10/matrix/*.png* .
```

Visualization of Contact Matrix: Plot an Interaction Matrix of Chr8



Visualization of Contact Matrix: Plot Multiple Matrices

##plot raw and normalized matrices side by side and show TAD domain panel:

```
python /u/home/galaxy/collaboratory/apps/HiCPlotter/  
HiCPlotter.py -f rawdata_500000.matrix  
rawdata_500000_iced.matrix -bed  
rawdata_500000_abs.bed -n raw iced -chr chr8 -o  
raw_iced_chr8 -tri 1 -r 500000 -hmc 5 -mm 10 -ptr 1 -pcd 1  
-pcdf hESC_domains_hg19.bed hESC_domains_hg19.bed
```

Visualization of Contact Matrix: Plot Multiple Matrices

