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:

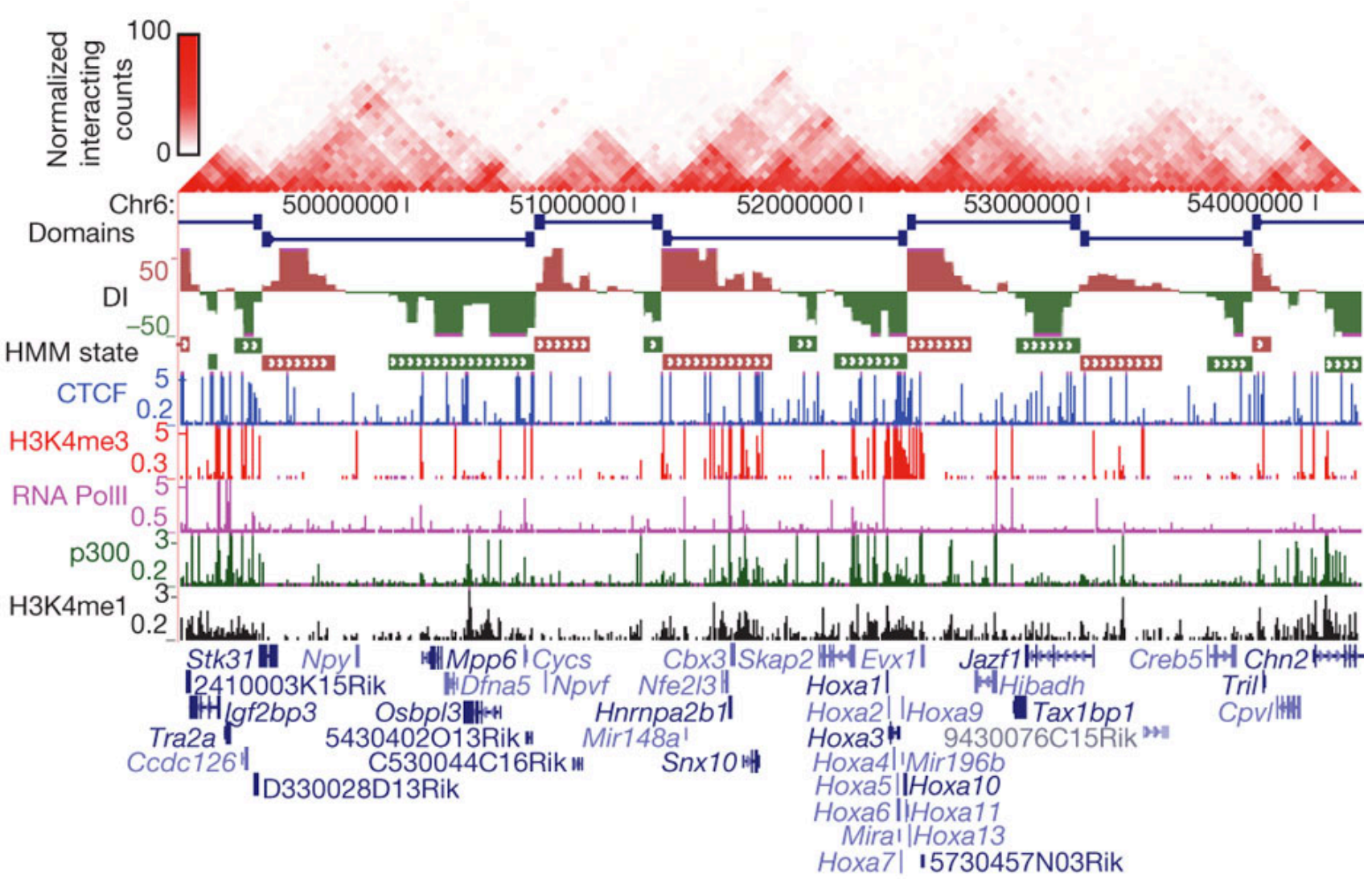
Aligning Hi-C library using HiC-Pro, continued.
Visualization of data using HiC-Plotter

Day 3:

Identification of significant chromatin interactions using GOTHiC.
Calculate directionality index using HiTC.
Overview of other useful Hi-C tools: hicup, HOMER

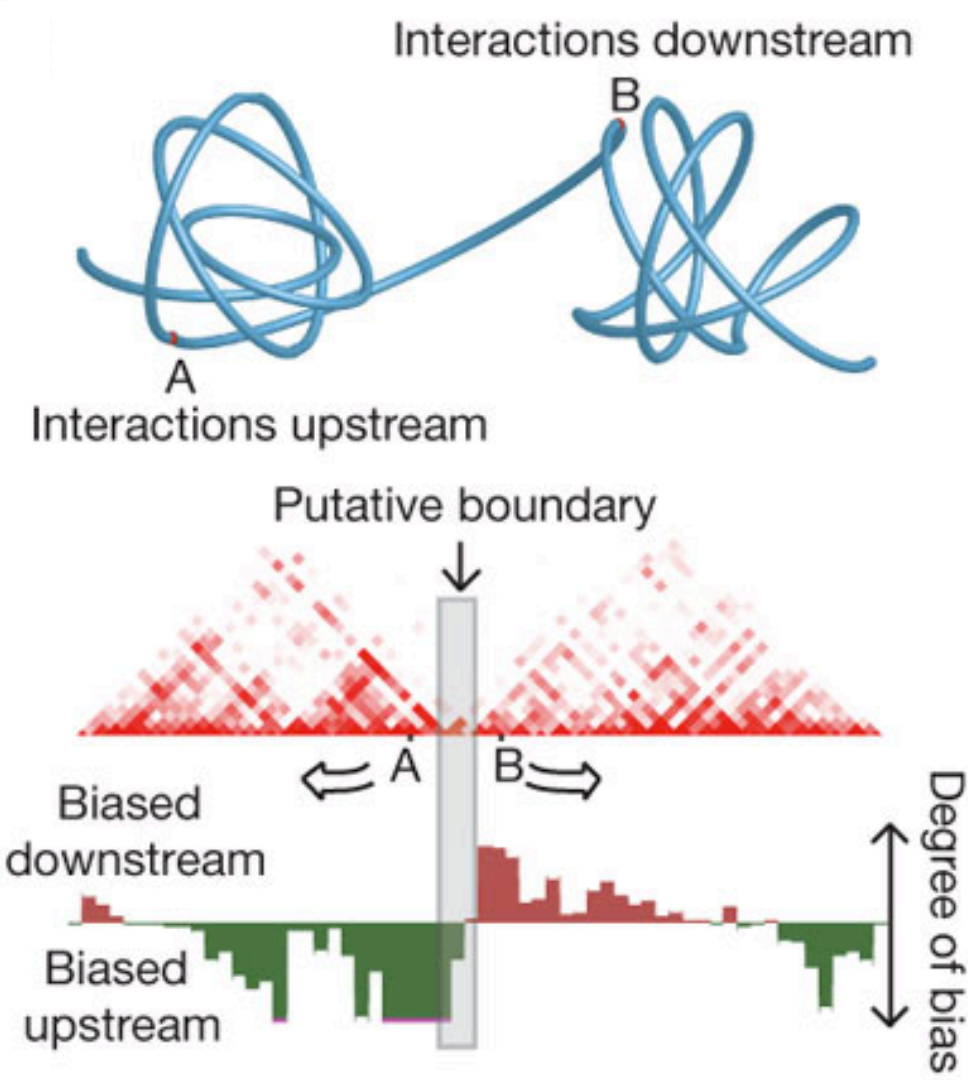
Day 3

Topological Associated Domains (TADs)



Dixon et at 2012

Identification of TADs by Directionality Index (DI)



Dixon et al Nature 2012

Hi-C Analysis Using Bioconductor Packages

##HiTC, for more information:
http://bioconductor.org/packages/release/bioc/html/HiTC.html

##GOTHiC
http://bioconductor.org/packages/release/bioc/html/GOTHiC.html

Hi-C Analysis Using HiTC: Calculating DI

```
##load R session:
    module load R/3.2.3
    R

##the following R packages and their dependencies will be used:
        BSgenome.Hsapiens.UCSC.hg19
        HiTC
        GOTHiC

##set working directory
    setwd("/u/scratch/f/flay/workshop10/matrix/")

##load these libraries:
    library('BSgenome.Hsapiens.UCSC.hg19')
    library('HiTC')
    library('GOTHiC')
```

Hi-C Analysis Using HiTC: Calculating DI

##to install these libraries:

```
source("http://bioconductor.org/biocLite.R")
biocLite("BSgenome.Hsapiens.UCSC.hg19")
biocLite(”HiTC")
biocLite(“GOTHiC")
```

##if prompted to install updates for other packages, type in no.

Hi-C Analysis Using HiTC: Calculating DI

##to import Hi-C data:

```
iced=importC("rawdata_500000_iced.matrix",
xgi.bed="rawdata_500000_abs.bed")
```

```
> iced=importC("rawdata_500000_iced.matrix", xgi.bed="rawdata_500000_abs.bed")
Loading Genomic intervals ...
Reading file ...
Convert 'C' file in HTCexp object(s)
Creating chr1-chr1 Contact Map ...
Creating chr2-chr1 Contact Map ...
Creating chr3-chr1 Contact Map ...
Creating chr4-chr1 Contact Map ...
Creating chr5-chr1 Contact Map ...
Creating chr6-chr1 Contact Map ...
Creating chr7-chr1 Contact Map ...
Creating chr8-chr1 Contact Map ...
Creating chr9-chr1 Contact Map ...
Creating chr10-chr1 Contact Map ...
Creating chr11-chr1 Contact Map ...
Creating chr12-chr1 Contact Map ...
Creating chr13-chr1 Contact Map ...
Creating chr14-chr1 Contact Map ...
```

Hi-C Analysis Using HiTC: Calculating DI

##summary of interactions in chr8:
detail(iced$chr8chr8)

```
> detail(iced$chr8chr8)
HTC object
Focus on genomic region [chr8:1-146364022]
CIS Interaction Map
Matrix of Interaction data: [293-293]
Binned data - window size = 500000
293 genome intervals
Total Reads =  3705556
Number of Interactions =  41170
Median Frequency =  10.85954
Sparsity =  0.48
```

Hi-C Analysis Using HiTC: Calculating DI

##let's focus on a region in the genome.

chr = extractRegion(iced$chr8chr8, chr="chr8", from=1, to=140e6)

##plot this region:

pdf("chr8.pdf")
plot(chr)
dev.off()

```
> chr = extractRegion(iced$chr8chr8, chr="chr8", from=1, to=140e6)
> pdf("chr8.pdf")
> plot(chr)
minrange= 1.598  - maxrange= 617.595
> dev.off()
null device
          1
```

Hi-C Analysis Using HiTC: Calculating DI

##calculate DI:

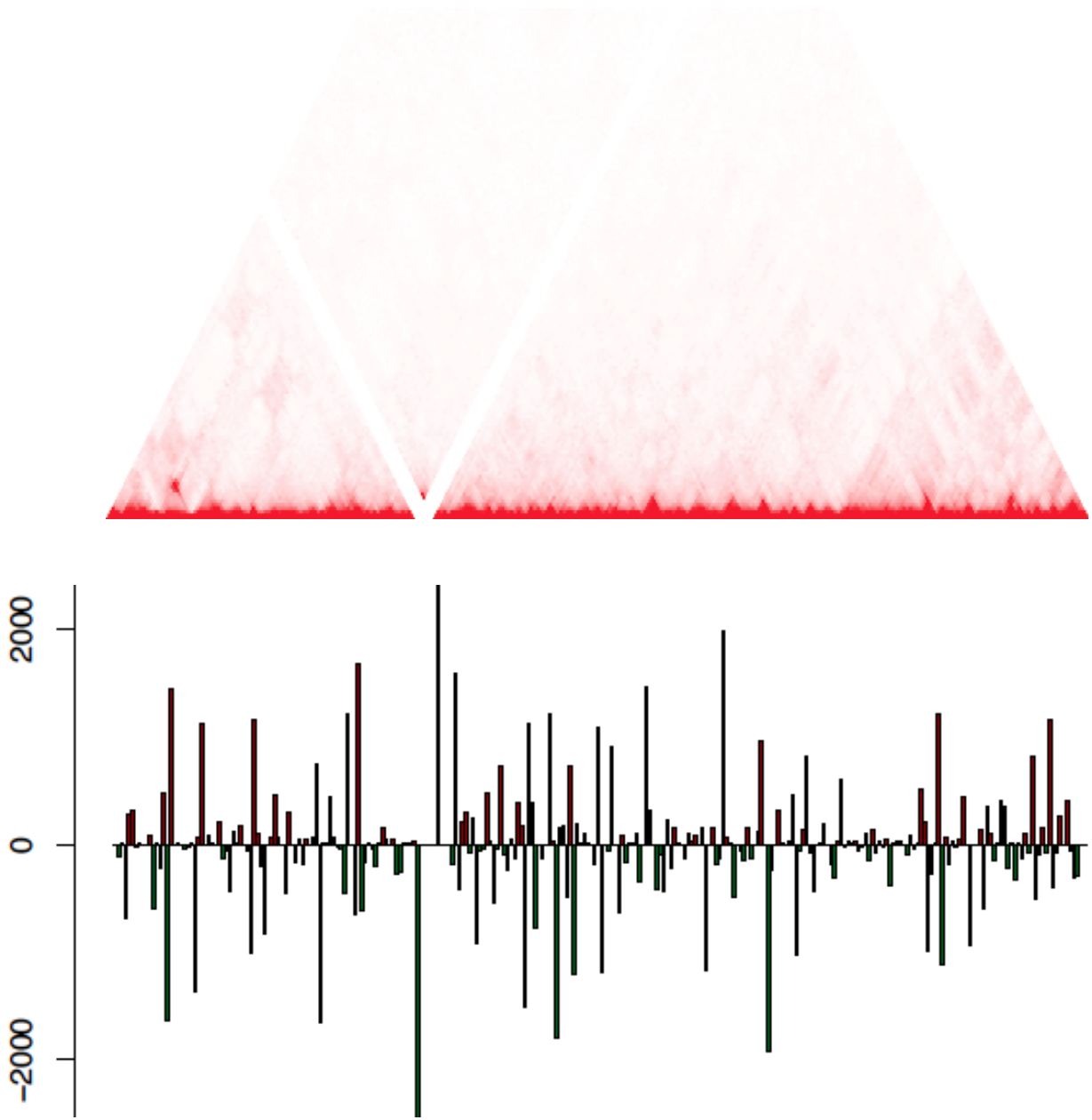
```
di=directionalityIndex(chr)
```

##plot DI:

```
pdf("chr_di.pdf")
barplot(di, col=ifelse(di>0, "darkred", "darkgreen"))
dev.off()
```

```
> di=directionalityIndex(chr)
> pdf("chr_di.pdf")
> barplot(di, col=ifelse(di>0,"darkred","darkgreen"))
> dev.off()
null device
          1
```

Hi-C Analysis Using HiTC: Visualize DI



Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##need to convert .bam file into GOTHiC readable format

cd /u/scratch/f/flay/workshop10/output/bowtie_results/bwt2/rawdata

```
[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
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##we will use tools from the hicup the software:
http://www.bioinformatics.babraham.ac.uk/projects/hicup/

##prepare hicup-compatible digested genome

hicup_digester

```
[flay@n2002 rawdata]$ hicup_digester

HiCUP homepage: www.bioinformatics.babraham.ac.uk/projects/hicup

The 'hicup_digester' script creates a reference genome, cut with a
specified restriction enzyme

SYNOPSIS

hicup_digester [OPTIONS]... -config [CONFIGURATION FILE]
hicup_digester [OPTIONS]... [FASTA FILES]...

FUNCTION

The HiCUP pipeline removes Hi-C artefacts, requiring a reference digested
genome. HiCUP Digester identifies the cut sites in FASTA files. The script
prints the results to file for subsequent processing by HiCUP Filter.

The names of the files to be processed and the digestion parameters may be
passed to the script by a configuration file or command line arguments.

COMMAND LINE OPTIONS

--re1           Restriction enzyme used to digest the genome (the enzyme that forms the
                ligation junction) e.g. A^GATCT,BglII.
                Some Hi-C protocols may use two enzymes at this stage.
                To specify two enzymes: -1 A^GATCT,BglII:A^AGCTT,HindIII.
--re2           To specify a restriction enzyme instead of sonication to shorten
                di-tags. This restriction site does NOT form a Hi-C ligation
                junction. 2 e.g. AG^CT,AluI. Typically the sonication
                protocol is followed.
--config        Specify the name of the optional configuration file
--genome        Name of the genome to be digested (not the path to the genome file
                or files, but the genome name to include in the output file)
--help          Print program help and exit
--outdir        Specify the directory to which the output files should be
                written
--quiet         Suppress all progress reports
--version       Print the program version and exit
--zip           Print the results to a gzip file

Full instructions on running the pipeline can be found at:
www.bioinformatics.babraham.ac.uk/projects/hicup

Steven Wingett, Babraham Institute, Cambridge, UK (steven.wingett@babraham.ac.uk)
[flay@n2002 rawdata]$
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##run hicup_digester for MboI

hicup_digester --genome hg19 --re1 ^GATC,MboI /u/scratch/f/flay/workshop10/genome/hg19_rCRSchrm.fa

```
[flay@n2002 rawdata]$ hicup_digester --genome hg19 --re1 ^GATC,MboI /u/scratch/f/flay/workshop10/genome/hg19_rCRSchrm.fa
HiCUP Digester (version 0.5.8)
Digesting files
Digesting '/u/scratch/f/flay/workshop10/genome/hg19_rCRSchrm.fa'
Digestion complete
```

##output file

```
[flay@n2002 rawdata]$ ls -lh
total 361M
-rw-r--r-- 1 flay matteop 308M May  5 21:52 Digest_hg19_MboI_None_21-50-09_05-05-2016.txt
-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
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##convert .bam file into GOTHiC format

/u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2gothic

```
[flay@n2002 rawdata]$
[flay@n2002 rawdata]$ /u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2gothic

HiCUP homepage: www.bioinformatics.babraham.ac.uk/projects/hicup

The hicup2gothic script converts HiCUP BAM/SAM files to a format compatible
with GOTHiC (GOTHiC Hi-C pages found at:
http://master.bioconductor.org/packages/release/bioc/html/GOTHiC.html).

SYNOPSIS

hicup2gothic [OPTIONS]
hicup2gothic [SAM/BAM FILES]...

FUNCTION

HiCUP generates SAM/BAM files of mapped, filtered paired-end reads
constituting the sequenced valid Hi-C di-tags. These may then be analysed by a
variety of specialised tools, but before this is possible the datasets will
need parsing into an appropriate format.

The hicup2gothic script converts HiCUP BAM/SAM files to a format compatible with
GOTHiC i.e. 4 columns, with reads on separate lines:

Column1: read ID
Column2: SAM flag
Column3: chromosome name
Column4: position

COMMAND LINE OPTIONS

--help         Print help message and exit
--version      Print the program version and exit

Full instructions on running the pipeline can be found at:
www.bioinformatics.babraham.ac.uk/projects/hicup

Steven Wingett, Babraham Institute, Cambridge, UK (steven.wingett@babraham.ac.uk)
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##convert .bam file into GOTHiC format

module load samtools/1.2

/u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2gothic test_dataset_hg19.bwt2pairs.bam

```
[flay@n2002 rawdata]$ /u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2gothic test_dataset_hg19.bwt2pairs.bam
Processing test_dataset_hg19.bwt2pairs.bam
Processing complete
```

##output file

```
[flay@n2002 rawdata]$ head test_dataset_hg19.bwt2pairs.bam.gothic
SRR071233.317   67      chr1    221618744
SRR071233.317   131     chr18   71915472
SRR071233.633   115     chr8    85211571
SRR071233.633   179     chr14   49035220
SRR071233.791   67      chr6    134957612
SRR071233.791   131     chr8    43087371
SRR071233.949   67      chr6    55376994
SRR071233.949   131     chr6    78497150
SRR071233.1107  83      chr1    39590414
SRR071233.1107  163     chr10   23824123
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

```
##load R
    library('GOTHiC')
    library('BSgenome.Hsapiens.UCSC.hg19')

##calculate binomial

binom=GOTHiChicup("test_dataset_hg19.bwt2pairs.bam.gothic",
sampleName="test_gothic", res=500000,
restrictionFile="Digest_hg19_MboI_None_21-50-09_05-05-2016.txt",
cistrans="all")

##turn off graphic setting after run to generate the summary plot
dev.off()
```

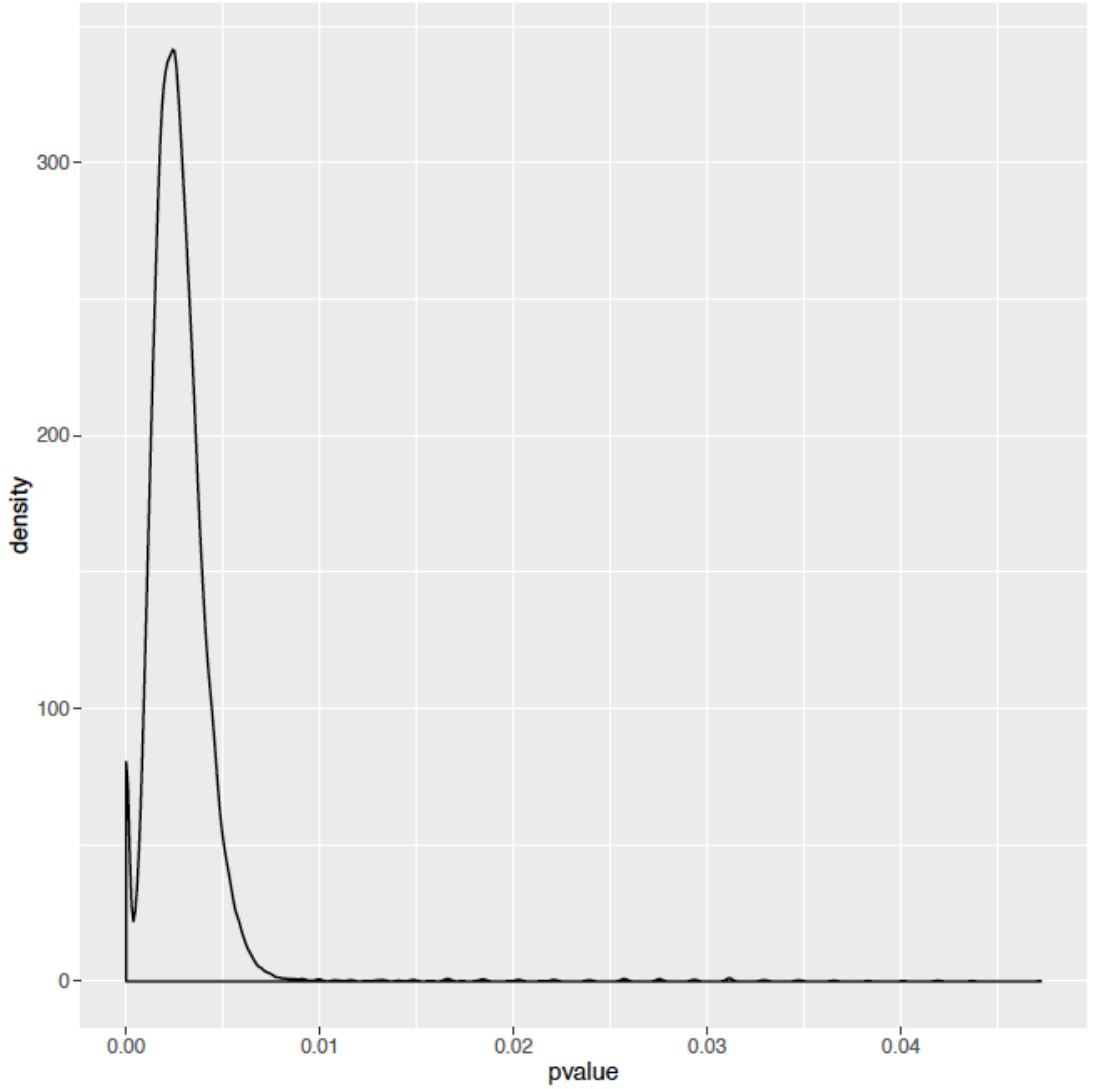
Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##binomial test result assumes biases affect both ends independently
##output is data frame

```
> head(binom)
       chr1 locus1  chr2    locus2 relCoverage1 relCoverage2  probability
106939 chr1  5e+05 chr11  48500000  0.000236039 0.0002221543 1.048951e-07
69221  chr1  5e+05 chr16  26000000  0.000236039 0.0002360390 1.114510e-07
120611 chr1  5e+05 chr18  58500000  0.000236039 0.0003054622 1.442308e-07
102065 chr1  5e+05  chr2  35000000  0.000236039 0.0001110772 5.244755e-08
108835 chr1  5e+05  chr2 224500000  0.000236039 0.0001943850 9.178321e-08
101191 chr1  5e+05  chr2 239500000  0.000236039 0.0002638083 1.245629e-07
          expected readCount      pvalue qvalue logObservedOverExpected
106939 0.003777377         1 0.003770252      1                 8.048399
69221  0.004013463         1 0.004005419      1                 7.960937
120611 0.005193894         1 0.005180431      1                 7.588968
102065 0.001888689         1 0.001886906      1                 9.048399
108835 0.003305205         1 0.003299749      1                 8.241045
101191 0.004485635         1 0.004475589      1                 7.800472
```

Hi-C Analysis Using GOTHiC: Identifying Significant Interactions

##p-value distribution



Contents lists available at ScienceDirect

Methods

journal homepage: www.elsevier.com/locate/ymeth

The Hitchhiker's guide to Hi-C analysis: Practical guidelines

Bryan R. Lajoie *, Job Dekker *, Noam Kaplan *

Program in Systems Biology, Department of Biochemistry and Molecular Pharmacology, University of Massachusetts Medical School, 368 Plantation Street, Worcester, MA 01605-0103, USA

ARTICLE INFO

Article history:
Received 5 August 2014
Received in revised form 28 October 2014
Accepted 30 October 2014
Available online 6 November 2014

Keywords:
Chromosome conformation capture
Deep sequencing
Chromatin structure
Bioinformatics
Hi-C

ABSTRACT

Over the last decade, development and application of a set of molecular genomic approaches based on the chromosome conformation capture method (3C), combined with increasingly powerful imaging approaches, have enabled high resolution and genome-wide analysis of the spatial organization of chromosomes. The aim of this paper is to provide guidelines for analyzing and interpreting data obtained with genome-wide 3C methods such as Hi-C and 3C-seq that rely on deep sequencing to detect and quantify pairwise chromatin interactions.

OPTIONAL SEGMENTS

Hi-C Analysis Using HOMER

##Detailed documentation can be found in
http://homer.salk.edu/homer/interactions/

HOMER

Software for motif discovery and next-gen sequencing analysis

---

Analyzing Hi-C genome-wide interaction data

HOMER contains several programs and analysis routines to facilitate the analysis of Hi-C data. Hi-C couples chromosome conformation capture (3C) with deep sequencing to reveal regions of genomic DNA that are in close spatial proximity in the nucleus. Hi-C has emerged as a powerful technique to understand how the genome is packaged in cells to control gene expression. Unlike ChIP-PET, 5C, or 4C, Hi-C is unbiased. While HOMER can be jury-rigged to gleam information from other 3C-sequencing based methods, it has been specifically tailored Hi-C analysis.

Specialized Hi-C programs in HOMER

HOMER has several specialized programs for Hi-C analysis. Each is covered in the tutorials below:

makeTagDirectory - special paired-end operations for making HOMER-style tag directories and filtering options for Hi-C

analyzeHiC - primary analysis program - generates interaction matrices, normalization, identification of significant interactions, clustering of domains, generates Circos plots (most of the following programs use this one internally)

runHiCpca.pl - automated PCA analysis on Hi-C data to identify "compartments"
getHiCcorrDiff.pl - calculates the difference in correlation profiles between two Hi-C experiments
findHiCCompartments.pl - find continuous or differential regions from PCA/corrDiff results that describe what compartment regions of DNA belong to (name changed - was called *getDomains.pl*)

findHiCInteractionsByChr.pl - helps automate the finding of high-resolution intra-chromosomal interactions
annotateInteractions.pl - program for re-analysis of significant interactions, such as relating them to ChIP-Seq peaks

SIMA.pl - Novel tool to boost sensitivity by pooling features together when performing interaction calculations

Hi-C Analysis Using HOMER

##convert HiC-Pro output to homer-compatible file

/u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2homer test_dataset_hg19.bwt2pairs.bam

```
[flay@n2001 homer]$ /u/home/galaxy/collaboratory/apps/hicup_v0.5.8/Conversion/hicup2homer test_dataset_hg19.bwt2pairs.bam
Processing test_dataset_hg19.bwt2pairs.bam
Processing complete
```

##homer format

```
[flay@n2001 homer]$ head test_dataset_hg19.bwt2pairs.bam.homer
1       chr1    221618744       +       chr18   71915472        +
2       chr8    85211571        -       chr14   49035220        -
3       chr6    134957612       +       chr8    43087371        +
4       chr6    55376994        +       chr6    78497150        +
5       chr1    39590414        -       chr10   23824123        +
6       chr11   84184431        -       chr6    23517254        +
7       chr11   78705346        +       chr13   36482601        -
8       chr6    67518942        +       chr6    67519165        -
9       chr16   22391788        -       chr3    107989046       -
10      chr12   66328972        +       chr12   76839007        -
```

Hi-C Analysis Using HOMER

##load homer

module load homer/4.7

##make tag directory

makeTagDirectory tag/ -format HiCsummary test_dataset_hg19.bwt2pairs.bam.homer

```
[flay@n2001 homer]$ makeTagDirectory tag/ -format HiCsummary test_dataset_hg19.bwt2pairs.bam.homer
        Will parse file: test_dataset_hg19.bwt2pairs.bam.homer

        Creating directory: tag/ and removing existing *.tags.tsv

        Reading alignment file test_dataset_hg19.bwt2pairs.bam.homer

        Optimizing tag files...
        Estimated genome size = 3094528916
        Estimated average read density = 0.000039 per bp
        Total Tags = 121970.0
        Total Positions = 121940
        Average tag length = 1.0
        Median tags per position = 1 (ideal: 1)
        Average tags per position = 1.000
        Fragment Length Estimate: 186
        Peak Width Estimate: 269
        Autocorrelation quality control metrics:
                Same strand fold enrichment: 2.6
                Diff strand fold enrichment: 9.6
                Same / Diff fold enrichment: 0.3

                Guessing sample is ChIP-Seq - uneven enrichment between same strand and
                different strands - may have problems such as clonal amplification.
```

Hi-C Analysis Using HOMER

```
[flay@n2001 homer]$ analyzeHiC

        Usage: analyzeHiC <PE tag directory> [options]
                ...

        Resolution Options:
                -res <#> (Resolution of matrix in bp or use "-res site" [see below], default: 10000000)
                -superRes <#> (size of region to count tags, i.e. make twice the res, default: same as res)

        Region Options:
                -chr <name> (create matrix on this chromosome, default: whole genome)
                -start <#> (start matrix at this position, default:0)
                -end <#> (end matrix at this position, default: no limit)
                -pos chrN:xxxxxx-yyyyyy (UCSC formatted position instead of -chr/-start/-end)
                -chr2 <name>, -start2 <#>, -end2 <#>, or -pos2 (Use these positions on the
                        y-axis of the matrix.  Otherwise the matrix will be sysmetric)
                -p <peak file> (specify regions to make matrix, unbalanced, use -p2 <peak file>)
                -vsGenome (normally makes a square matrix, this will force 2nd set of peaks to be the genome)
                -chopify (divide up peaks into regions the size of the resolution, default: use peak midpoints)
                -fixed (do not scale the size of the peaks to that of the resolution)
                -relative (use with -maxDist <#>, outputs diagonal of matrix up to maxDistance)
                -pout <filename> (output peaks used for analysis as a peak file, -pout2 <file> for 2nd set)

        Interaction Matrix Options:
                -norm (normalize by dividing each position by expected counts [log ratio], default)
                -zscoreNorm (normalize log ratio by distance dependent stddev, add "-nolog" to return linear values)
                -logp (output log p-values)
                -simpleNorm (Only normalize based on total interactions per location [log ratio], not distance)
                -raw (report raw interaction counts)
                -expected (report expected interaction counts based on average profile)
                -rawAndExpected <filename for expectd matrix> (raw counts sent to stdout)
                -corr (report Pearson's correlation coeff, use "-corrIters <#>" to recursively calculate)
                        -corrDepth <#> (merge regions in correlation so that minimum # expected tags per data point)
                -o <filename> (Output file name, default: sent to stdout)
                -cluster (cluster regions, uses "-o" to name cdt/gtr files, default: out.cdt)
                -clusterFixed (clusters adjacent regions, good for linear domains)
                -std <#> (# of std deviations from mean interactions per region to exclude, default:4)
                -min <#> (minimum fraction of average read depth to include in analysis, default: 0.2)
                -override (Allow very large matrices to be created... be carful using this)
```

Hi-C Analysis Using HOMER

```
Interaction Matrix Options:
        -norm (normalize by dividing each position by expected counts [log ratio], default)
        -zscoreNorm (normalize log ratio by distance dependent stddev, add "-nolog" to return linear values)
        -logp (output log p-values)
        -simpleNorm (Only normalize based on total interactions per location [log ratio], not distance)
        -raw (report raw interaction counts)
        -expected (report expected interaction counts based on average profile)
        -rawAndExpected <filename for expectd matrix> (raw counts sent to stdout)
        -corr (report Pearson's correlation coeff, use "-corrIters <#>" to recursively calculate)
                -corrDepth <#> (merge regions in correlation so that minimum # expected tags per data point)
        -o <filename> (Output file name, default: sent to stdout)
        -cluster (cluster regions, uses "-o" to name cdt/gtr files, default: out.cdt)
        -clusterFixed (clusters adjacent regions, good for linear domains)
        -std <#> (# of std deviations from mean interactions per region to exclude, default:4)
        -min <#> (minimum fraction of average read depth to include in analysis, default: 0.2)
        -override (Allow very large matrices to be created... be carful using this)

Background Options:
        -fullModel (perform exhaustive background model calculations, default: approxmiate)
        -quickModel (perform approximate background model calculations, default: approxmiate)
        -force (force the creation of a fresh genome-wide background model)
        -bgonly (quit after creating the background model)
        -createModel <custom bg model output file> (Create custome bg from regions specified, i.e. -p/-pos)
        -model <custom bg model input file> (Use Custom background model, -modelBg for -ped)
        -randomize <bgmodel> <# reads> (and the output is a PE tag file, initail PE tag directory not used
                Use makeTagDirectory ... -t outputfile to create the new directory)
Non-matrix stuff:
        -nomatrix (skip matrix creation - use if more than 100,000 loci)
        -interactions <filename> (output interactions, add "-center" to adjust pos to avg of reads)
        -pvalue <#> (p-value cutoff for interactions, default 0.001)
        -zscore <#> (z-score cutoff for interactions, default 1.0)
        -minDist <#> (Minimum interaction distance, default: resolution/2)
        -maxDist <#> (Maximum interaction distance, default: -1=none)
        -boundary <#> (score boundaries at the given scale i.e. 100,000)

Making Histograms:
        -hist <#> (create a histogram matrix around peak positions, # is the resolution)
        -size <#> (size of region in histogram, default = 100 * resolution)
```

Hi-C Analysis Using HOMER

```
        Comparing HiC experiments:
                -ped <background PE tag directory>

        Creating BED file to view with Wash U Epigenome Browser:
                -washu (Both matrix and interaction outputs will be in WashH BED format)

        Creating Circos Diagrams:
                -circos <filename prefix> (creates circos files with the given prefix)
                -d <tag directory 1> [tag directory 2] ... (will plot tag densities with circos)
                -b <peak/BED file> (similar to visiualization of genes/-g, but no labels)
                -g <gene location file> (shows gene locations)

        Given Interaction Analysis Mode (no matrix is produced):
                -i <interaction input file> (for analyzing specific sets of interactions)
                -iraw <output BED filename> (output raw reads from interactions, or -irawtags <file>)
                -4C <output BED file> (outputs tags interacting with specified regions)
                -peakStats <output BED file prefix> (outputs several UCSC bed/bedGraph files with stats)
                -cpu <#> (number of CPUs to use, default: 1)

[flay@n2001 homer]$
```

Hi-C Analysis Using HOMER

##create background model

analyzeHiC tag/ -res 1000000 -bgonly

```
[flay@n2001 homer]$ analyzeHiC tag/ -res 1000000 -bgonly
        Genome Size=3094132969.0
        No HiC background model found for 1000000 bp resolution. Creating...
        Generating Background using -fullModel
        Genome Size=3094132969.0
        Calculating PE Tag Coverage:.............................................
        Avg interactions per peak = 39.0 +/- 17.6

        Finding Interactions to average into expected profile (-fullModel)...
                chr1
                chr2
                chr3
                chr4
                chr5
                chr6
                chr7
                chr8
                chr9
                chr10
                chr11
                chr12
                chr13
                chr14
                chr15
                chr16
                chr17
                chr18
                chr19
                chr20
                chr21
                chr22
                chrX
                chrY
                chrM
```

Hi-C Analysis Using HOMER

##create “normalized” interaction matrix for the whole genome
##large memory required

analyzeHiC tag/ -res 1000000 -norm -override > test_whole_genome_matrix.txt

```
[flay@n2001 homer]$ analyzeHiC tag/ -res 1000000 -norm -override > test_whole_genome_matrix.txt
        Genome Size=3094132969.0
        Found HiC background model for 1000000 bp resolution (-force to overwrite)
        Total regions in background model=3127
        Total Interactions: 121970.0
        Model accuracy: 1.000000
        Using model to derive PE Tag densities

        Average interaction count in regions = 39.0 +/- 17.6
                chr1
                chr2
                chr3
```

##create normalized interaction matrix for one chromosome

analyzeHiC tag/ -res 1000000 -norm -chr chr8 > test_chr8_matrix.txt

Hi-C Analysis Using HOMER

##run principal component analysis on Hi-C data:
runHiCpca.pl test_pca tag/ -res 1000000 -superRes 1000000 -genome hg19 -corrDepth 1

```
[flay@n2001 homer]$ runHiCpca.pl test_pca tag/ -res 1000000 -superRes 1000000 -genome hg19 -corrDepth 1
        Using TSS from hg19 to assign sign to active(+) vs. inactive(-) PC1 values
        Output files will start with: test_pca
        Analyzing HiC directory: tag/

        Will analyze chrs: chr10 chr11_gl000202_random chr11 chr12 chr13 chr14 chr15 chr16 chr17_gl000204_random chr17_gl000205_random chr17 chr18 chr19_gl0
00208_random chr19_gl000209_random chr19 chr1_gl000192_random chr1 chr20 chr21 chr22 chr2 chr3 chr4_gl000193_random chr4_gl000194_random chr4 chr5 chr6 chr7
_gl000195_random chr7 chr8 chr9_gl000198_random chr9_gl000199_random chr9 chrM chrUn_gl000211 chrUn_gl000212 chrUn_gl000214 chrUn_gl000216 chrUn_gl000218 ch
rUn_gl000219 chrUn_gl000220 chrUn_gl000224 chrUn_gl000225 chrUn_gl000229 chrUn_gl000231 chrUn_gl000232 chrUn_gl000234 chrUn_gl000235 chrUn_gl000240 chrUn_gl
000241 chrX chrY
        Genome Size=3094132969.0
        Found HiC background model for 1000000 bp resolution (-force to overwrite)
        Total regions in background model=3127
        Total Interactions: 121970.0
        Model accuracy: 1.000000
        Using model to derive PE Tag densities

        Average interaction count in regions = 43.1 +/- 14.0
                chr10

        Reporting linear (not log2) ratios
        Pooling regions during correlation calculation (need >1.0 expected reads)
        Calculating Correlation Matrix:
        |0%                                  |50%                                  100%|
        ================================================================================
        Printing Interaction Matrix: 136 x 136
        Genome Size=3094132969.0
        Found HiC background model for 1000000 bp resolution (-force to overwrite)
        Total regions in background model=3127
        Total Interactions: 121970.0
        Model accuracy: -nan
        Using model to derive PE Tag densities

        Average interaction count in regions = -nan +/- -nan

        Reporting linear (not log2) ratios
        Pooling regions during correlation calculation (need >1.0 expected reads)
        Calculating Correlation Matrix:
```

Hi-C Analysis Using HOMER

##output files of pca analysis

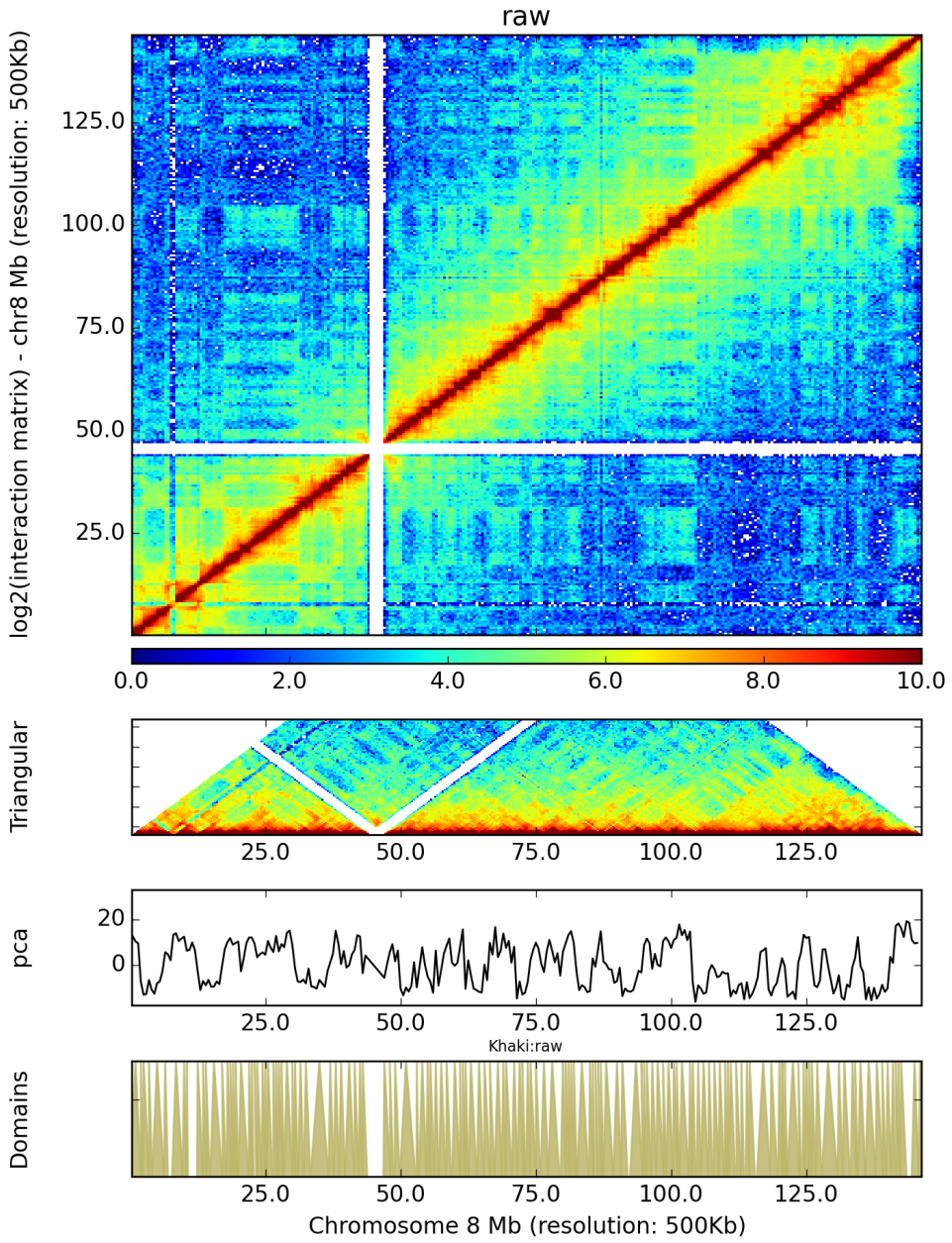
```
[flay@n2001 homer]$ head test_pca.PC1.bedGraph
track name="test_pca PC1"  yLineMark="0.0" alwaysZero=on maxHeightPixels=100:75:11 visibility=full viewLimits=-1:1 autoScale=on type=bedGraph
chr1    0       1000000 -1.92363327845944
chr1    1000000 2000000 2.07410292679
chr1    2000000 3000000 3.50406406578341
chr1    3000000 4000000 -2.96320736855501
chr1    4000000 5000000 -3.89707024999703
chr1    5000000 6000000 -6.61197525822796
chr1    6000000 7000000 8.19043488997856
chr1    7000000 8000000 4.36221791110957
chr1    8000000 9000000 2.80062742913176
[flay@n2001 homer]$
[flay@n2001 homer]$ head test_pca.PC1.txt
#peakID chr     start   end     strand  PC1
chr10-0 chr10   0       1000000 +       9.19494959346629
chr10-1000000   chr10   1000000 2000000 +       -8.21438401849641
chr10-2000000   chr10   2000000 3000000 +       -7.33338470493469
chr10-3000000   chr10   3000000 4000000 +       -4.18804701963403
chr10-4000000   chr10   4000000 5000000 +       8.55032702838669
chr10-5000000   chr10   5000000 6000000 +       5.91495112982676
chr10-6000000   chr10   6000000 7000000 +       -5.59221761634582
chr10-7000000   chr10   7000000 8000000 +       8.83036912629083
chr10-8000000   chr10   8000000 9000000 +       -7.04240184439858
```

Hi-C Analysis Using HOMER

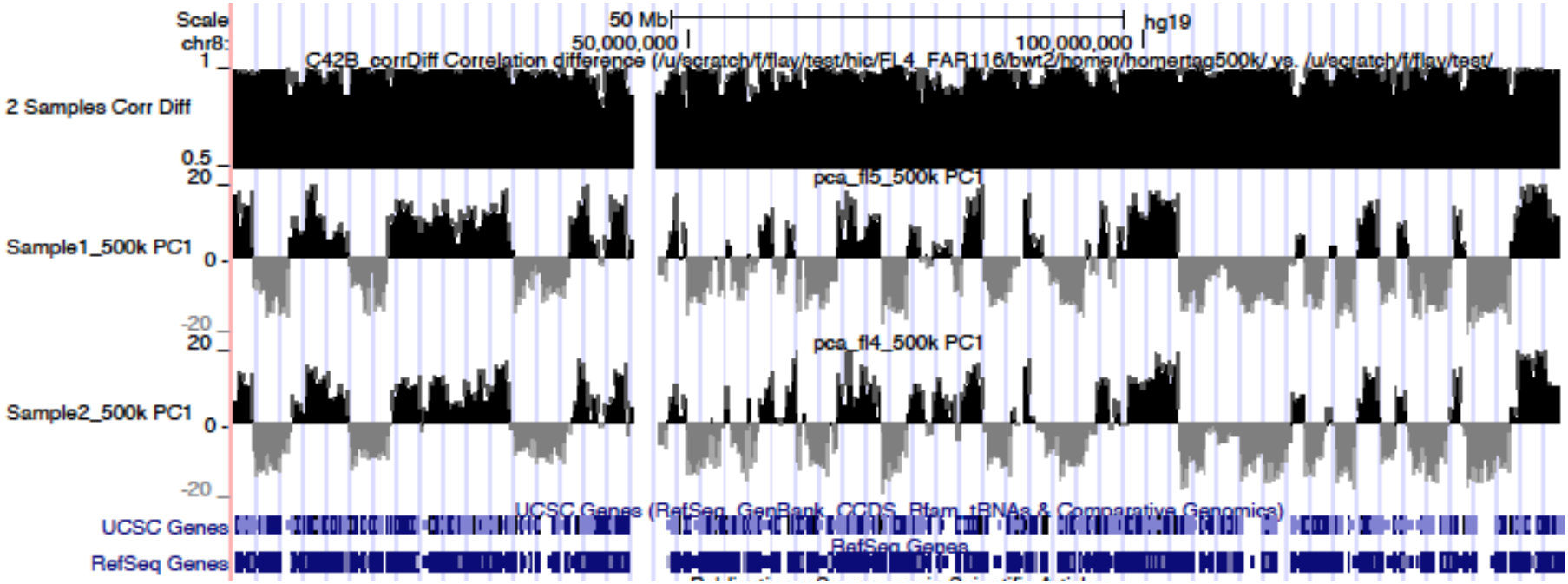
##visualize pca using HiCPlotter

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 -pcd 1 -pcdf hESC_domains_hg19.bed -hist pca_fl4_500k.PC1.bedGraph -hl pca -hm 30

Hi-C Analysis Using HOMER



Hi-C Analysis Using HOMER



Hi-C Analysis Using HOMER

##identify significant interactions

analyzeHiC tag/ -res 1000000 -superRes 1000000 -interactions test_significantInteractions.txt -nomatrix

Hi-C Analysis Using HOMER

```
[flay@n2178 homer]$ analyzeHiC tag/ -res 1000000 -superRes 1000000 -interactions test_significantInteractions.txt -nomatrix
        Genome Size=3094132969.0
        Found HiC background model for 1000000 bp resolution (-force to overwrite)
        Total regions in background model=3127
        Total Interactions: 121970.0
        Model accuracy: 1.000000
        Using model to derive PE Tag densities

        Average interaction count in regions = 39.0 +/- 17.6
                chr1
                chr2
                chr3
                chr4
                chr5
                chr6
                chr7
                chr8
                chr9
                chr10
                chr11
                chr12
                chr13
                chr14
                chr15
                chr16
                chr17
                chr18
                chr19
                chr20
                chr21
                chr22
                chrX
                chrY
                chrM
                chr17_gl000204_random
                chr17_gl000205_random
                chr19_gl000208_random
                chr1_gl000192_random
                chr4_gl000193_random
                chr7_gl000195_random
                chr9_gl000198_random
                chr9_gl000199_random
                chrUn_gl000211
                chrUn_gl000214
                chrUn_gl000218
                chrUn_gl000219
                chrUn_gl000220
                chrUn_gl000224
                chrUn_gl000225
                chrUn_gl000231
                chrUn_gl000232
                chrUn_gl000234
                chrUn_gl000241

        Regions with too many or too few reads: 2.00%
        Total Significant Interactions: 3
```