Workshop 6: DNA Methylation Analysis using Bisulfite Sequencing

Fides D Lay
UCLA
QCB Fellow
lay.fides@gmail.com

Workshop 6 Outline

Day 1:

Introduction to DNA methylation & WGBS
Quick review of linux, Hoffman2 and high-throughput sequencing glossary.
Aligning WGBS reads using bwa-meth

Day 2:

DNA methylation calling using Bis-SNP
Analysis of differentially methylated regions (DMRs) using metilene

Day 3:

Visualization of DNA methylation data
WGBS analysis using BS-Seeker2

Day 2

Methylation Calling using Bis-SNP

Liu *et al. Genome Biology* 2012, **13**:R61
http://genomebiology.com/2012/13/7/R61

METHOD Open Access

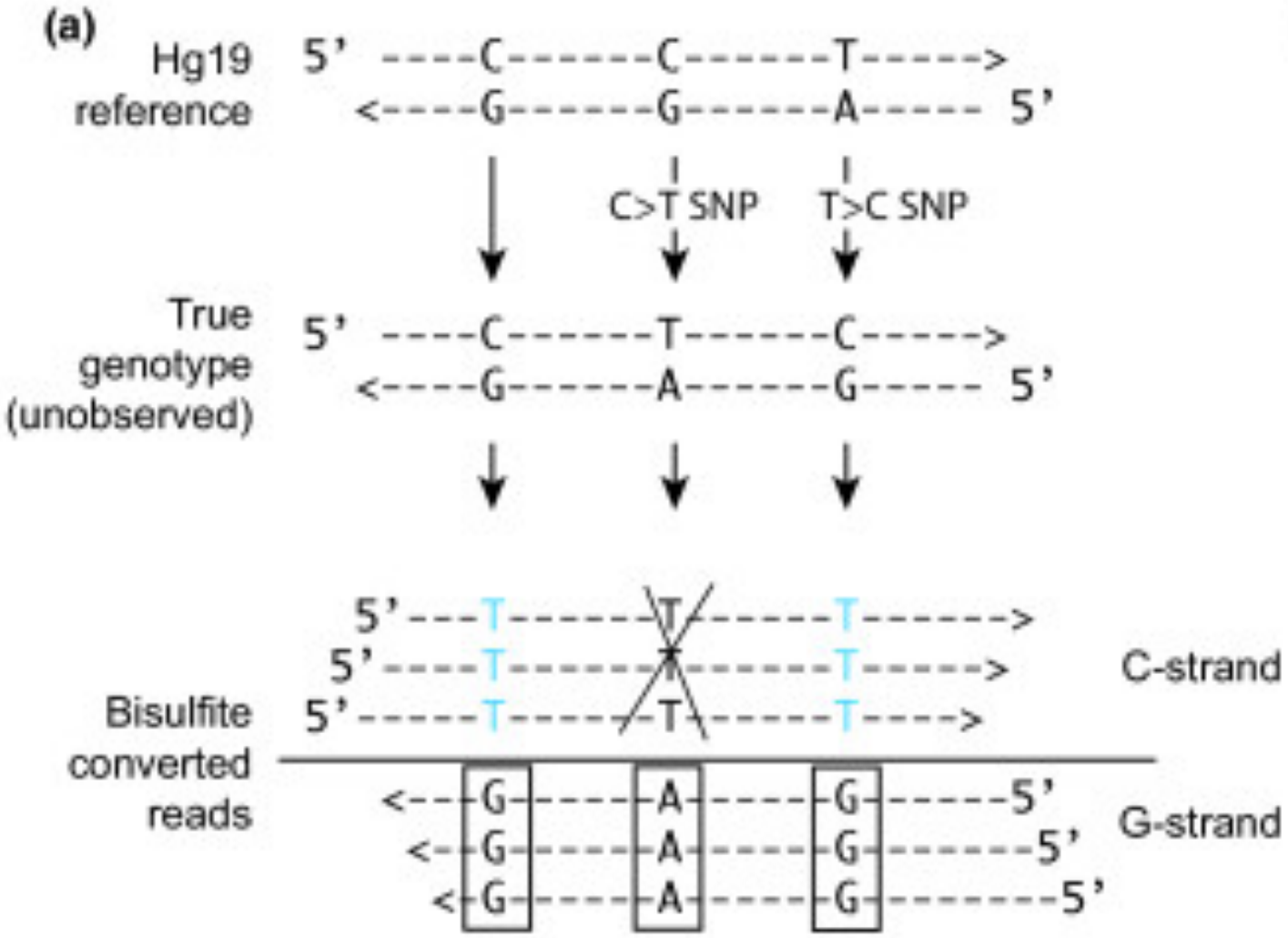
Bis-SNP: Combined DNA methylation and SNP calling for Bisulfite-seq data

Yaping Liu[1,2], Kimberly D Siegmund[3], Peter W Laird[1] and Benjamin P Berman[1,3*]

Abstract

Bisulfite treatment of DNA followed by high-throughput sequencing (Bisulfite-seq) is an important method for studying DNA methylation and epigenetic gene regulation, yet current software tools do not adequately address single nucleotide polymorphisms (SNPs). Identifying SNPs is important for accurate quantification of methylation levels and for identification of allele-specific epigenetic events such as imprinting. We have developed a model-based bisulfite SNP caller, Bis-SNP, that results in substantially better SNP calls than existing methods, thereby improving methylation estimates. At an average 30× genomic coverage, Bis-SNP correctly identified 96% of SNPs using the default high-stringency settings. The open-source package is available at http://epigenome.usc.edu/publicationdata/bissnp2011.

Methylation Calling using Bis-SNP



Liu et al 2012

Methylation Calling using Bis-SNP

Advantages:

- Simultaneous variant calling to determine DNA methylation and genomic variants
- Calling of cytosines present in the context of GCH as used in NOMe-seq assay

Disadvantages:

- Computationally intensive & slower
- Currently only compatible with Java6
- No alignment tool by itself

Methylation Calling using Bis-SNP

people.csail.mit.edu/dnaase/bissnp2011/

- Home
- SF Project Page
- SF Download Page
- Google group
- SVN Browser
- Useful Utilies & Files
- User Manual
- Quick Start & Test Dataset
- Step by step genotyping tutorial
- Variant Call Format
- GATK
- Picard tools
- Contact

Bis-SNP

------ A bisulfite space genotyper & methylation caller

Introduction

BisSNP is a package based on the Genome Analysis Toolkit (GATK) map-reduce framework for genotyping and accurate DNA methylation calling in bisulfite treated massively parallel sequencing (Bisulfite-seq, NOMe-seq, RRBS and any other bisulfite treated sequencing) with Illumina directional library protocol. It contains the following key features:

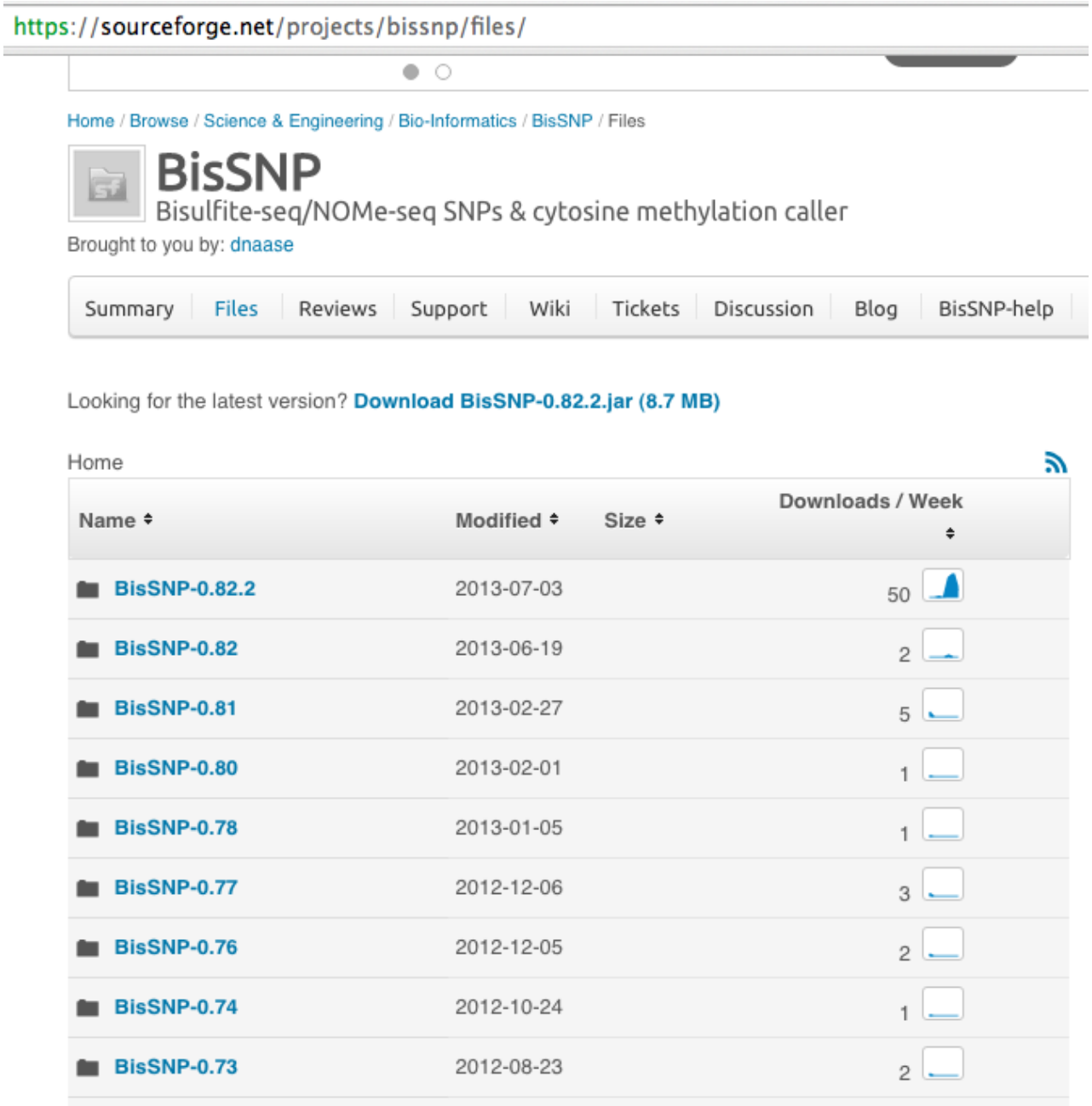
- Call and summarize methylation of any cytosine context provided (CpG, CHH, CHG, GCH et.al.);
- Work for single end and paired-end data;
- Accurtae variant detection. Enable base quality recalibration and indel calling in bisulfite sequencing;
- Based on Java map-reduce framework, allow multi-thread computing. Cross-platform;
- Allow multiple output format, detailed VCF files, CpG haplotype reads file for mono-allelic methylation analysis, simplified bedGraph, wig and bed format for visualization in UCSC genome broswer and IGV browser.

BisSNP uses bayesian inference with locus specific methylation probabilities and bisulfite conversion rate of different cytosine context(not only CpG, CHH, CHG in Bisulfite-seq, but also GCH et.al. in other bisulfite treated sequencing) to determine genotypes and methylation levels simultaneously. Specificity and sensitivity has been validate by Illumina IM SNP array. In default threshold (Phred scale score > 20), it could detect 92.21% heterozygous SNPs with 0.14% false positive rate (90.88% sensitivity in C/T SNPs with 0.16% false positive rate, 98.51% sensitivity in non C/T SNPs with 0.16% false positive rate). Cytosine calling is not only based on reference context, so it could detect non-reference cytosine context for usage in epigenome wide association study.

BisSNP is hosted by SourceForge.net. The project page is here. The executable jar file are available from the download page. You can check out the latest source codes with:

svn checkout svn://svn.code.sf.net/p/bissnp/code/trunk bissnp-code

Methylation Calling using Bis-SNP



Methylation Calling using Bis-SNP

#bwa-meth v0.10 contains a wrapper for Bis-SNP

```
[flay@login3 ~]$ bwameth.py tabulate
usage:
    tabulate methylation from bwameth.py call
     [-h]
                                                    [--reference REFERENCE]
                                                    [-t THREADS]
                                                    [--dbsnp DBSNP]
                                                    [--prefix PREFIX]
                                                    [--trim TRIM]
                                                    [--map-q MAP_Q]
                                                    --bissnp BISSNP
                                                    [--region REGION]
                                                    [--format FORMAT]
                                                    [--context {all,CG,CG-strict} | --nome]
                                                    bams [bams ...]

    tabulate methylation from bwameth.py call
```

Methylation Calling using Bis-SNP

```
#bwa-meth contains a wrapper for Bis-SNP for easier use
#load dependencies module
    module load java/1.6.0_23
    module load gatk
    module load python/2.7.3
    module load bwa
    module load samtools/1.2


python /u/home/galaxy/collaboratory/apps/bwa-meth-0.10/
bwameth.py tabulate --reference /u/scratch/f/flay/workshop6/
genome/hg19_rCRSchrm.fa --threads 5 --prefix N25_bissnp --
dbsnp /u/scratch/f/flay/workshop6/genome/dbsnp/
dbsnp_135.hg19.sort.vcf --bissnp /u/scratch/f/flay/workshop6/
tools/Bis-SNP/BisSNP-0.87.jar --context CG  N25.bam
```

Methylation Calling using Bis-SNP

```
[flay@n2066 raw]$ python /u/home/galaxy/collaboratory/apps/bwa-meth-0.10/bwameth.py tabulate --reference /u/scrat
ch/f/flay/workshop6/genome/hg19_rCRSchrm.fa --threads 5 --prefix N25_bissnp --dbsnp /u/scratch/f/flay/workshop6/g
enome/dbsnp/dbsnp_135.hg19.sort.vcf --bissnp /u/scratch/f/flay/workshop6/tools/Bis-SNP/BisSNP-0.87.jar --context
CG N25.bam
    java -Xmx24g -jar /u/scratch/f/flay/workshop6/tools/Bis-SNP/BisSNP-0.87.jar \
        -R /u/scratch/f/flay/workshop6/genome/hg19_rCRSchrm.fa \
        -I N25.bam \
        -T BisulfiteGenotyper \
        --trim_5_end_bp 2 \
        --trim_3_end_bp 2 \
        -vfn1 N25_bissnp.meth.vcf -vfn2 N25_bissnp.snp.vcf \
        --non_directional_protocol \
        -mbq 12 \
        -minConv 0 \
        -toCoverage 1000 \
        -mmq 25 --dbsnp /u/scratch/f/flay/workshop6/genome/dbsnp/dbsnp_135.hg19.sort.vcf  \
        -nt 5
N25_bissnp.meth.vcf
```

Methylation Calling using Bis-SNP: Output Files

```
[flay@n2176 raw]$ ls -lh
total 8.3M
-rw-r--r-- 1 flay matteop  188 Jun 16 22:23 bwameth_align.sh
-rw-r--r-- 1 flay matteop 906K Jun 16 22:17 N25.bam
-rw-r--r-- 1 flay matteop 1.6M Jun 16 22:17 N25.bam.bai
-rw-r--r-- 1 flay matteop  489 Jun 16 22:52 N25_bissnp.command.sh
-rw-r--r-- 1 flay matteop 1.3M Jun 16 22:52 N25_bissnp.meth.vcf
-rw-r--r-- 1 flay matteop 1.1K Jun 16 22:52 N25_bissnp.meth.vcf.MethySummarizeList.txt
-rw-r--r-- 1 flay matteop 128K Jun 16 22:52 N25_bissnp.N25_R.meth.bed
-rw-r--r-- 1 flay matteop  14K Jun 16 22:52 N25_bissnp.snp.vcf
-rw-r--r-- 1 flay matteop 1.3M Jun 16 22:04 N25_R1.fastq
-rw-r--r-- 1 flay matteop 1.3M Jun 16 22:04 N25_R2.fastq
```

```
[flay@n2176 raw]$ head N25_bissnp.N25_R.meth.bed
#chrom  start   start   pct     cs      ts
chr1    2422460 2422460 100.0   1       0
chr1    2422478 2422478 100.0   1       0
chr1    2422520 2422520 100.0   1       0
chr1    1256567 1256567 100.0   1       0
chr1    1256694 1256694 100.0   1       0
chr1    1256704 1256704 100.0   1       0
chr1    6467474 6467474 100.0   1       0
chr1    7049773 7049773 100.0   1       0
chr1    9585829 9585829 0.0     0       1
```

Assessing Library Quality: Bisulfite Conversion Rate

#open the summary file

view N25_bissnp.meth.vcf.MethySummarizeList.txt

```
BisSNP version                                  BisSNP-0.87
Visited bases                                   646762
Callable bases                                  614014
Confidently called bases                        602155
Visited bases in Cpg island                     0
Callable bases in Cpg island                    0
Confidently called bases in Cpg island          0
% callable bases of all loci                    94.937
% confidently called bases of all loci          93.103
% confidently called bases of callable loci     98.069
Actual calls made                               602155
Average good reads coverage in all visited loci                         0.7
Average good reads coverage in callable position                        0.7
Average good reads coverage in all position of Cpg island                       NaN
Average good reads coverage in callable position of Cpg island                        NaN
##Methylation summary in total:
C:      112249  3.254%
CH:     104644  0.549%
CG:     3901    76.339%
##Methylation summary in Read Group:    N25_R
C:      112249  3.254%
CG:     3901    76.339%
CH:     104644  0.549%
```

Quick Recap: Features and Variation of DNA Methylation

Rakyan et al 2011

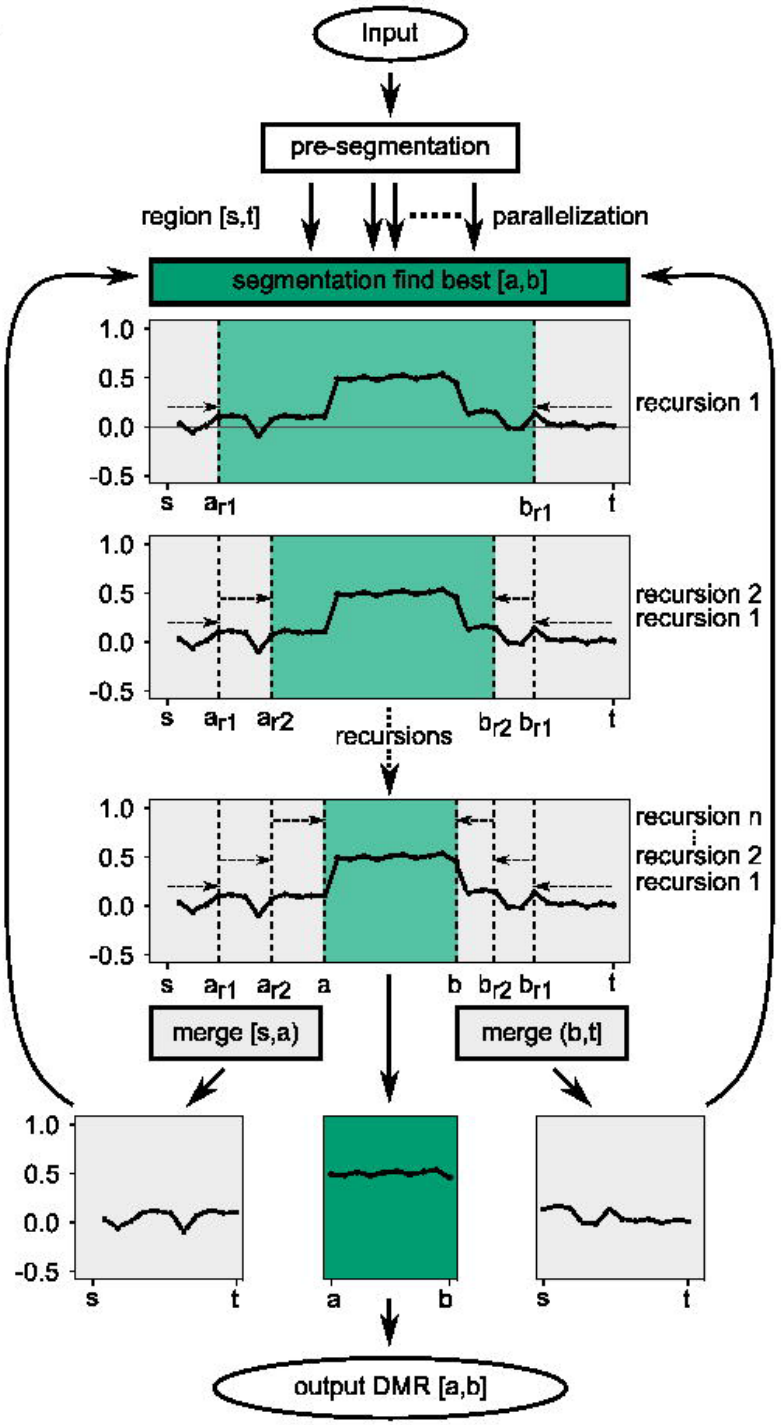
Identifying Differentially Methylated Regions (DMRs) Using metilene

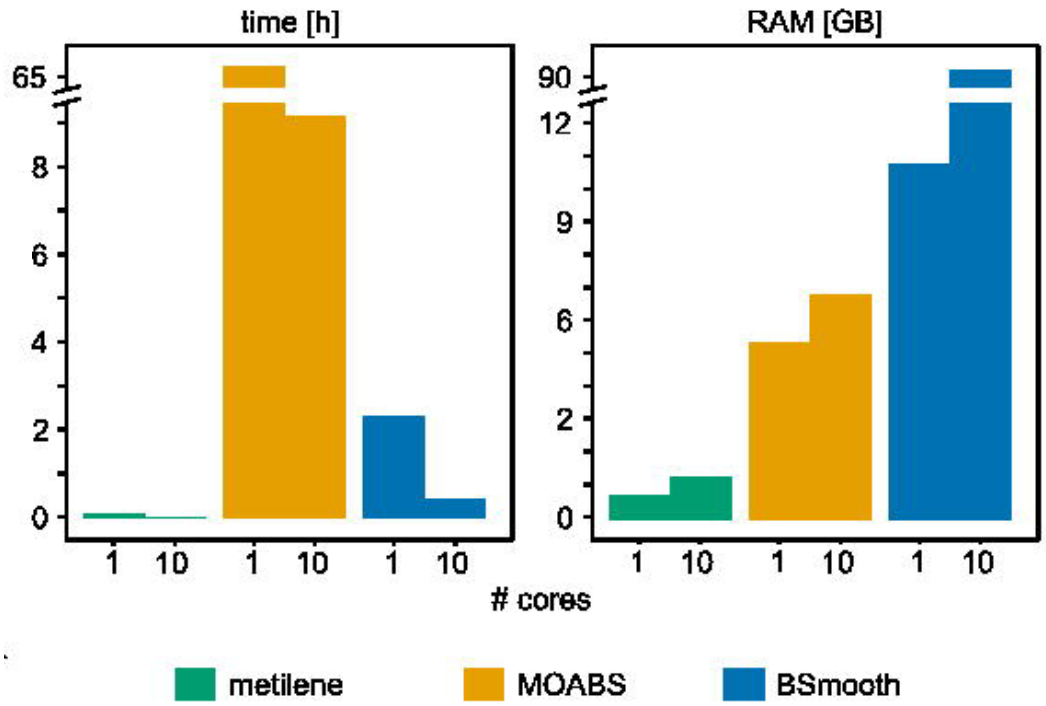
metilene: Fast and sensitive calling of differentially methylated regions from bisulfite sequencing data

Frank Jühling, Helene Kretzmer, Stephan H. Bernhart, et al.

Genome Res. published online December 2, 2015
Access the most recent version at doi:10.1101/gr.196394.115

metilene Workflow





Juhling et al, 2015

Identifying Differentially Methylated Regions (DMRs) Using metilene

- Installation
  - ✓Installing software
  - ✓Adding to environment ($PATH)
- Preparing bwa-meth output files for metilene
  - ✓Convert bed to bedgraph
  - ✓Generate metilene input files
- DMR analysis
  - ✓Running metilene
  - ✓Filtering

Identifying Differentially Methylated Regions (DMRs) Using metilene

- **Installation**
  - ✓**Installing software**
  - ✓**Adding to environment ($PATH)**
- Preparing bwa-meth output files for metilene
  - ✓Convert bed to bedgraph
  - ✓Generate metilene input files
- DMR analysis
  - ✓Running metilene
  - ✓Filtering

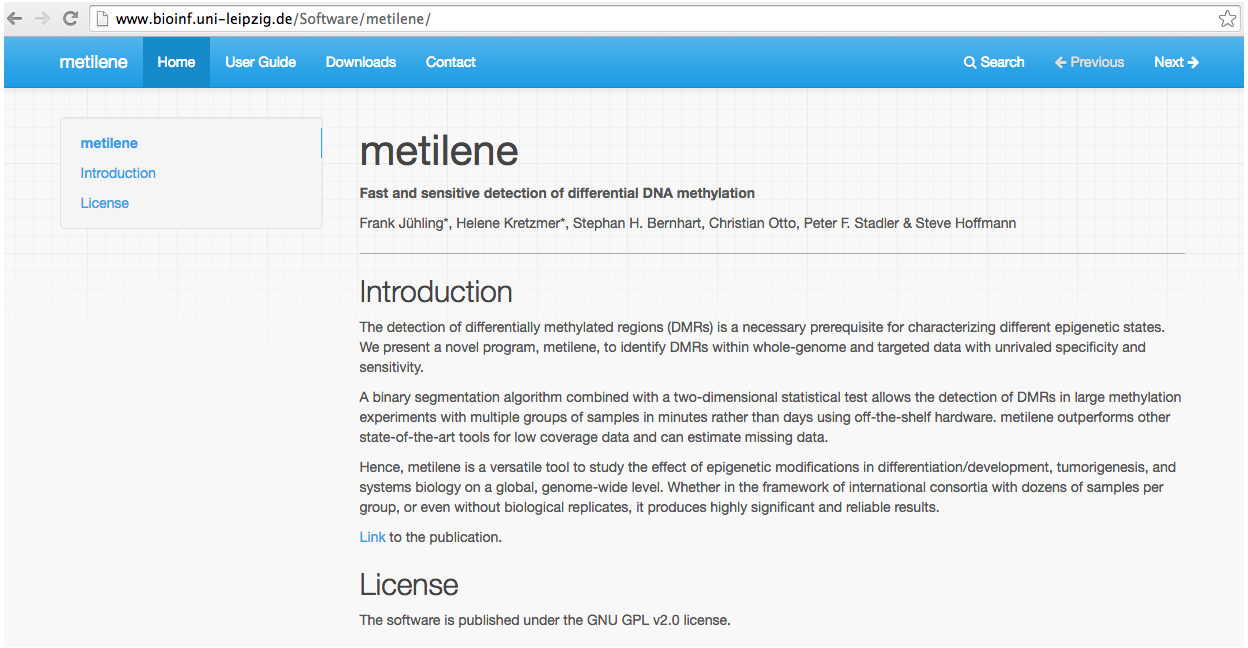
Install metilene

#change directory to where metilene will be installed

```
mkdir $HOME/software
cd $HOME/software
```

#download the latest version of metilene from
http://www.bioinf.uni-leipzig.de/Software/metilene/

wget http://www.bioinf.uni-leipzig.de/Software/metilene/metilene_v02-5.tar.gz



Install metilene

#unpack file

tar –xvzf metilene_v02-5.tar.gz

#enter the directory

cd metilene_v0.2-5

#always read the notes and instructions before doing anything

view README

#to install

make

Install metilene

#add the directory containing metilene into your $PATH by editing your ~/.bash_profile

view ~/.bash_profile

#to insert, type i and copy and paste the directory

i

#to save changes

:wq!

source ~/.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/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:/u/home/galaxy/collaboratory/apps/HiCPlotter:/u/home/galaxy/collaboratory/apps/FastQC:/u/home/f/flay/.local/bin

#export PATH
#export LD_LIBRARY_PATH=$HOME/.local/lib/:$LD_LIBRARY_PATH
#export PYTHONPATH=/u/home/f/flay/.local/lib/python2.7/site-packages/:$PYTHONPATH
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
~
```

Install metilene

```
[flay@n2192 metilene_v0.2-5]$ metilene
metilene: no source file provided.
usage: metilene [-M <n>] [-m <n>] [-d <n>] [-t <n>] [-f <n>] [-a <string>] [-b <string>] [-B <string>] [-X <n>] [-Y <n>] [-v <n>]  DataInputfile
  metilene - a tool for fast and sensitive detection of differential DNA methylation

DataInputFile           needs to be SORTED for chromosomes and genomic positions
 -M, --maxdist <n>      maximum distance (default:300)
 -m, --mincpgs <n>      minimum cpgs (default:10)
 -d, --minMethDiff <n>  minimum mean methylation difference (default:0.100000)
 -t, --threads <n>      number of threads (default:1)
 -f, --mode <n>         number of method: 1: de-novo, 2: pre-defined regions, 3: DMCs (default:1)
 -a, --groupA <string>  name of group A (default:"g1")
 -b, --groupB <string>  name of group B (default:"g2")
 -B, --bed <string>     bed-file for mode 2 containing pre-defined regions; needs to be SORTED equally to the DataInputFile (default:none)
 -X, --minNoA <n>       minimal number of values in group A (default:-1)
 -Y, --minNoB <n>       minimal number of values in group B (default:-1)
 -v, --valley <n>       valley filter (0.0 - 1.0) (default:0.700000)
 [VERSION]
  0.2-5
 [BUGS]
  Please report bugs to [frank,steve]@bioinf.uni-leipzig.de
 [REFERENCES]
  Implemented by Frank Juehling and Steve Hoffmann
  2015-2016 Bioinformatik Leipzig
```

Identifying Differentially Methylated Regions (DMRs) Using metilene

- Installation
  - ✓Installing software
  - ✓Adding to environment ($PATH)
- **Preparing bwa-meth output files for metilene**
  - **✓Convert bed to bedgraph**
  - **✓Generate metilene input files**
- DMR analysis
  - ✓Running metilene
  - ✓Filtering

Formatting Files To Input into metilene

#Convert .bed file to .bedgraph

```
sed '1d' fileName.bed | awk '{print $1 "\t" $2 "\t" $3 "\t" $4}' > fileName.bedgraph
```

Remove first line from file

Print the first four columns

```
[flay@n2176 raw]$ sed '1d' N25_bissnp.N25_R.meth.bed | awk '{print $1 "\t" $2 "\t" $3 "\t" $4}' > N25_bissnp.N25_R.meth.bedgraph
[flay@n2176 raw]$
[flay@n2176 raw]$ head N25_bissnp.N25_R.meth.bedgraph
chr1    2422460 2422460 100.0
chr1    2422478 2422478 100.0
chr1    2422520 2422520 100.0
chr1    1256567 1256567 100.0
chr1    1256694 1256694 100.0
chr1    1256704 1256704 100.0
chr1    6467474 6467474 100.0
chr1    7049773 7049773 100.0
chr1    9585829 9585829 0.0
chr1    9585872 9585872 100.0
```

Formatting Files To Input into metilene

```
#load dependency module
    module load bedtools/2.23.0

#sort all .bedgraph files
    sortBed -i fileName.bedgraph > fileName_sort.bedgraph

#Generate input files

perl /u/home/f/flay/software/metilene/metilene_input.pl -in1
file1.bedgraph file2.bedgraph -in2 name1.bedgraph
name2.bedgraph -h1 file -h2 name -o
metilene_file_name.input
```

Formatting Files To Input into metilene

#Today's input files are already in
/u/scratch/f/flay/workshop6/data/metilene/metilene_BL_FL.input

#To see what the input file look like:
head metilene_BL_FL.input

chr	pos	file_1	file_2	file_...	name_1	name_2	name_...

Group 1 with n samples

Group 2 with n samples

Identifying Differentially Methylated Regions (DMRs) Using metilene

- Installation
  - ✓Installing software
  - ✓Adding to environment ($PATH)
- **Preparing bwa-meth output files for metilene**
  - ✓**Convert bed to bedgraph**
  - ✓**Generate metilene input files**
- **DMR analysis**
  - ✓**Running metilene**
  - ✓**Filtering**

metilene Options

```
[flay@n2192 metilene_v0.2-5]$ metilene
metilene: no source file provided.
usage: metilene [-M <n>] [-m <n>] [-d <n>] [-t <n>] [-f <n>] [-a <string>] [-b <string>] [-B <string>] [-X <n>] [-Y <n>] [-v <n>]  DataInputfile
  metilene - a tool for fast and sensitive detection of differential DNA methylation

DataInputFile           needs to be SORTED for chromosomes and genomic positions
 -M, --maxdist <n>      maximum distance (default:300)
 -m, --mincpgs <n>      minimum cpgs (default:10)
 -d, --minMethDiff <n>  minimum mean methylation difference (default:0.100000)
 -t, --threads <n>      number of threads (default:1)
 -f, --mode <n>         number of method: 1: de-novo, 2: pre-defined regions, 3: DMCs (default:1)
 -a, --groupA <string>  name of group A (default:"g1")
 -b, --groupB <string>  name of group B (default:"g2")
 -B, --bed <string>     bed-file for mode 2 containing pre-defined regions; needs to be SORTED equally to the DataInputFile (default:none)
 -X, --minNoA <n>       minimal number of values in group A (default:-1)
 -Y, --minNoB <n>       minimal number of values in group B (default:-1)
 -v, --valley <n>       valley filter (0.0 - 1.0) (default:0.700000)
 [VERSION]
  0.2-5
 [BUGS]
  Please report bugs to [frank,steve]@bioinf.uni-leipzig.de
 [REFERENCES]
  Implemented by Frank Juehling and Steve Hoffmann
  2015-2016 Bioinformatik Leipzig
```

Running metilene: Identify DMRs

#Run metilene on default mode:

metilene -a BL -b FL metilene_BL_FL.input | sort -V -k1,1 -k2,2n > metilene_BL_FL.output

chr	start	stop	q-value	Mean meth. diff.	#CpG	p(MWU)	p(2D KS)	Mean methylation group 1 (-a)	Mean methylation group 2 (-b)

#How many DMRs were identified at this stage?

wc -l metilene_BL_FL.output

1101

Running metilene: Filter DMRs

#Let's look at our DMRs more closely:

module load R

```
[flay@n6190 metilene]$ perl /u/home/f/flay/software/metilene_v0.2-5/metilene_output.pl

    usage:  perl metilene_output.pl  -q <query_file> [-o <path_prefix>] [-p  <number>] [-c <number>] [-d <number>] [-l <number>] [-a <string>] [-b <string>]

    [INPUT]     -q          path/filename of metilene DMRs
                -o          path/prefix of output files (default: input_path/)
                -p          maximum (<) adj. p-value (q-value) for output of significant DMRs (default: 0.05)
                -c          minimum (>=) cpgs (default:10)
                -d          minimum mean methylation difference (>=) (default:0.1)
                -l          minimum length of DMR [nt] (>=) (post-processing, default: 0)
                -a          name of group A (default:"g1")
                -b          name of group B (default:"g2")
```

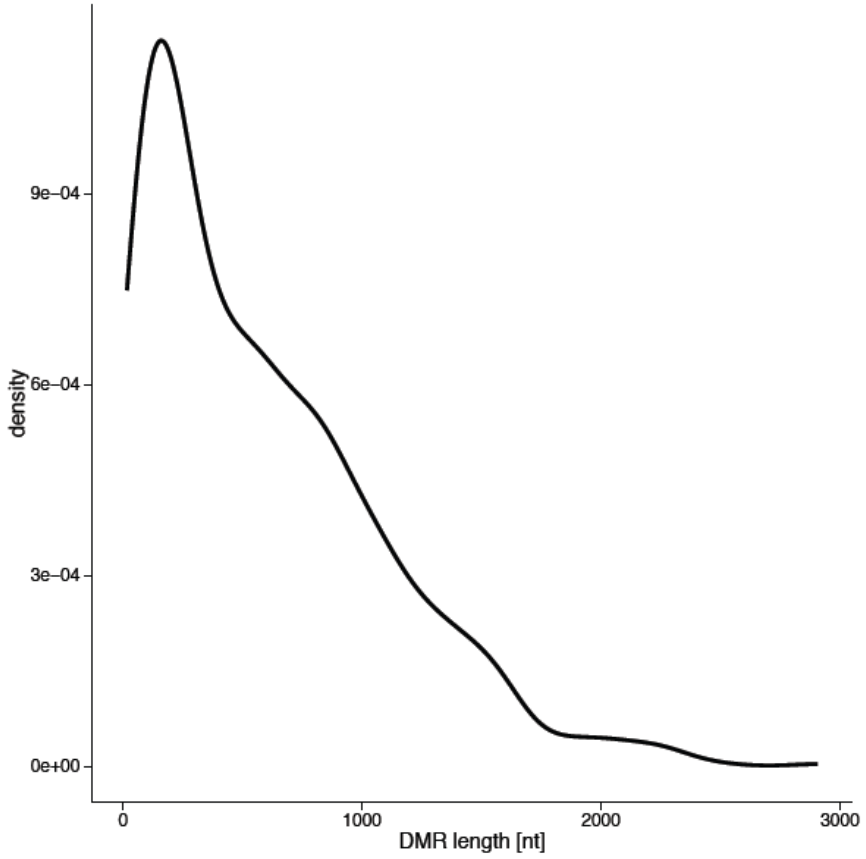
perl /u/home/f/flay/software/metilene_v0.2-5/metilene_output.pl -q metilene_BL_FL.output -o metilene_BL_FL.filter -a BL -b FL

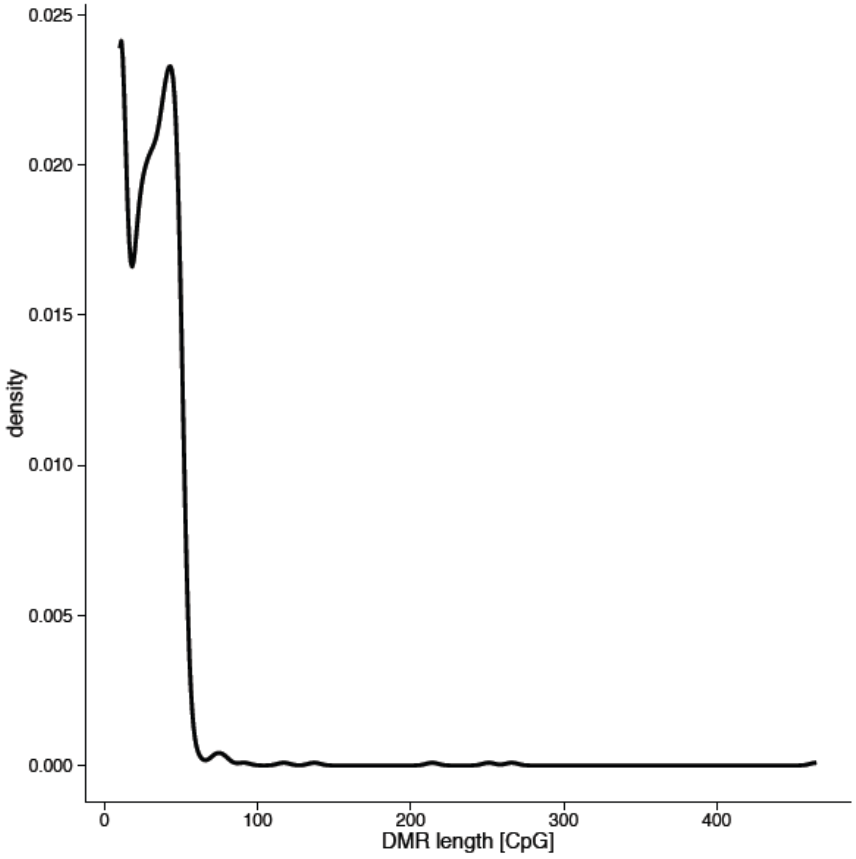
```
[flay@n6190 metilene]$ perl /u/home/f/flay/software/metilene_v0.2-5/metilene_output.pl -q metilene_BL_FL.output -o metilene_BL_FL.filter -a BL -b FL
[INFO]  Mon Mar 7, 10:29:8, 2016        Checking flags
[INFO]  Mon Mar 7, 10:29:8, 2016        Filter DMRs.
[INFO]  Mon Mar 7, 10:29:8, 2016        Wrote 1096 DMRs with adj. p-value<0.05, a minimum absolute difference>=0.1, a minimum length [CpG]>=10 and a minimum
 length [nt]>=0
[INFO]  Mon Mar 7, 10:29:8, 2016        Bedgraph file containing DMR difference: /u/scratch/f/flay/workshop6/data/metilene/metilene_BL_FL.filter_qval.0.05.b
edgraph
[INFO]  Mon Mar 7, 10:29:8, 2016        Plot DMR statistics.
Loading required package: methods
null device
          1
```

Running metilene: Output

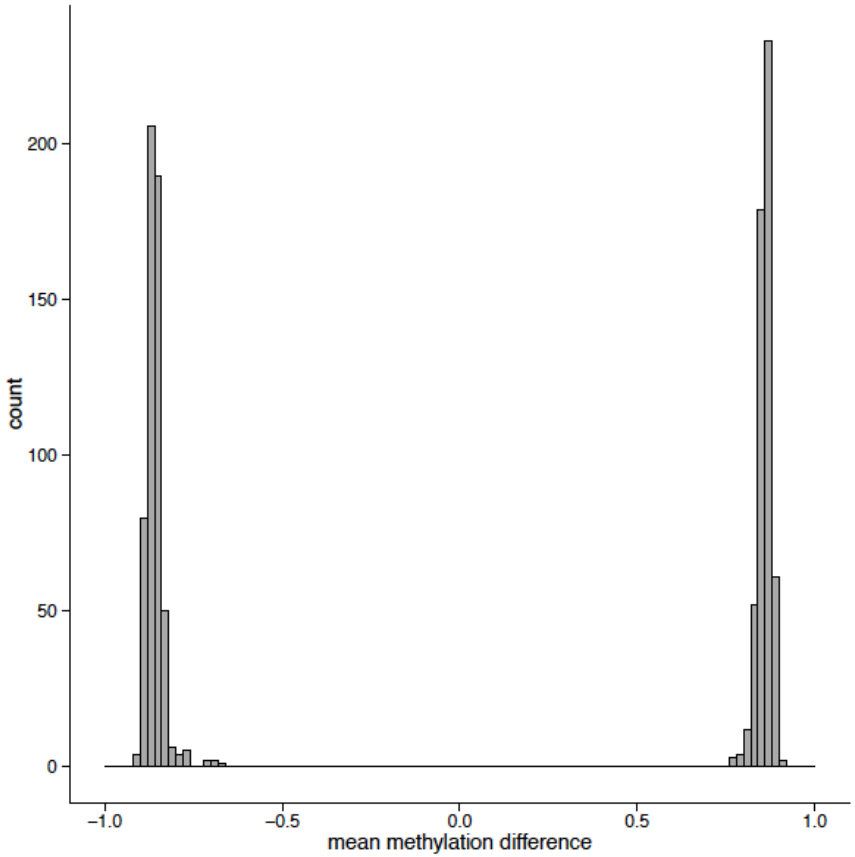
```
[flay@n6190 metilene]$ ls -lh
total 810M
-rw-r--r-- 1 flay matteop  37K Mar  7 10:29 metilene_BL_FL.filter_qval.0.05.bedgraph
-rw-r--r-- 1 flay matteop  67K Mar  7 10:29 metilene_BL_FL.filter_qval.0.05.out
-rw-r--r-- 1 flay matteop 193K Mar  7 10:29 metilene_BL_FL.filter_qval.0.05.pdf
-rw-r--r-- 1 flay matteop 718M Mar  6 12:09 metilene_BL_FL.input
-rw-r--r-- 1 flay matteop  84K Mar  7 10:18 metilene_BL_FL.output
```

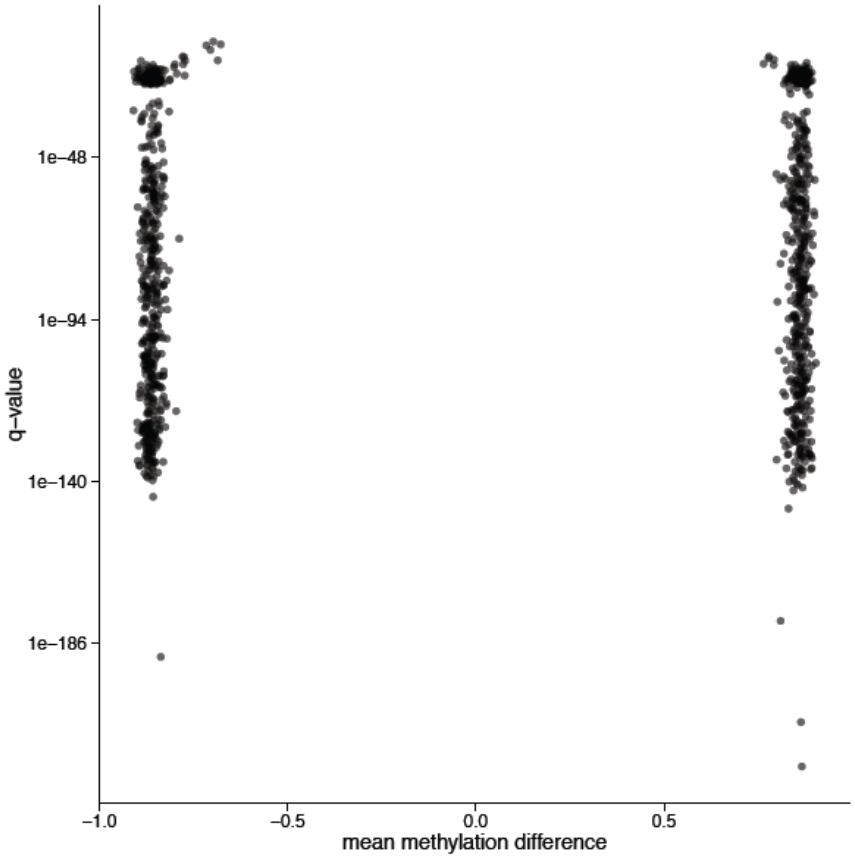
DMR Statistics



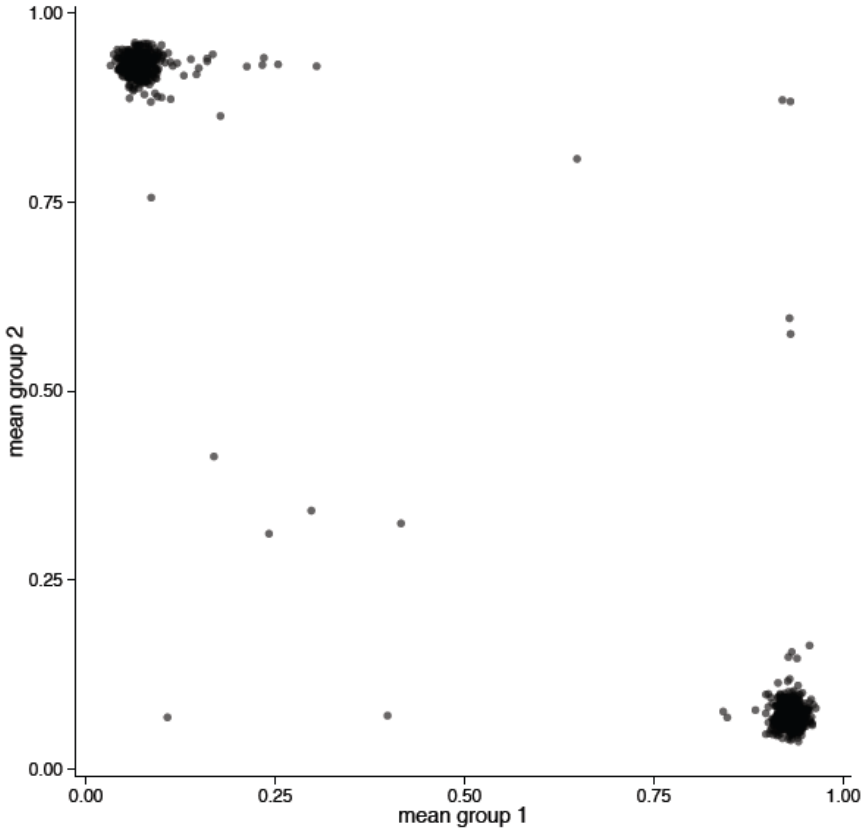


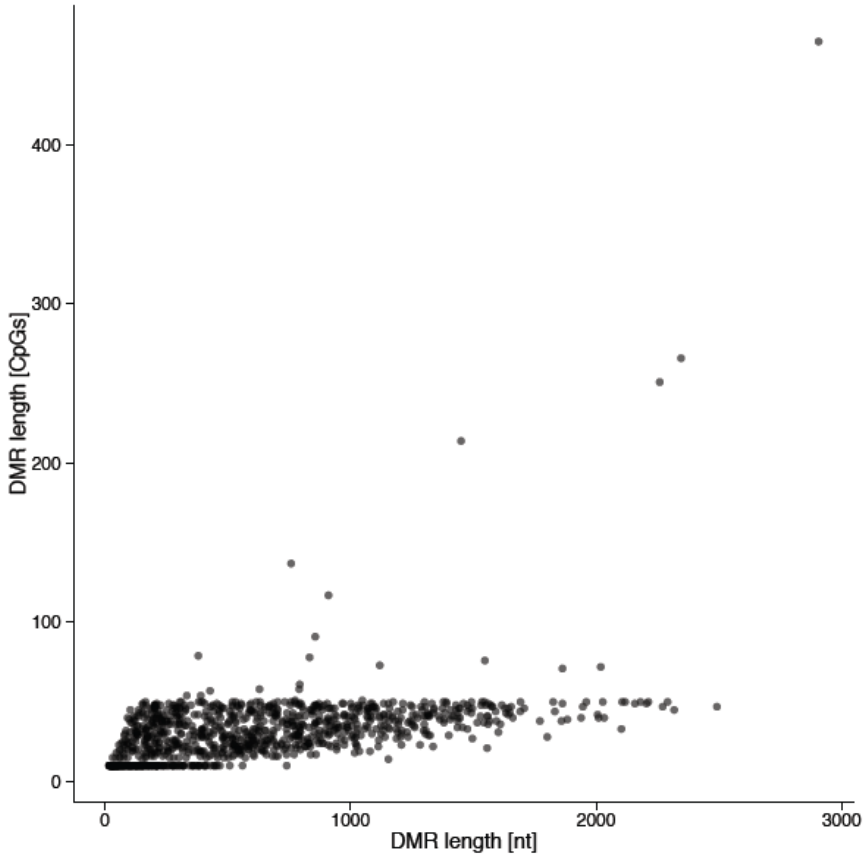
DMR Statistics





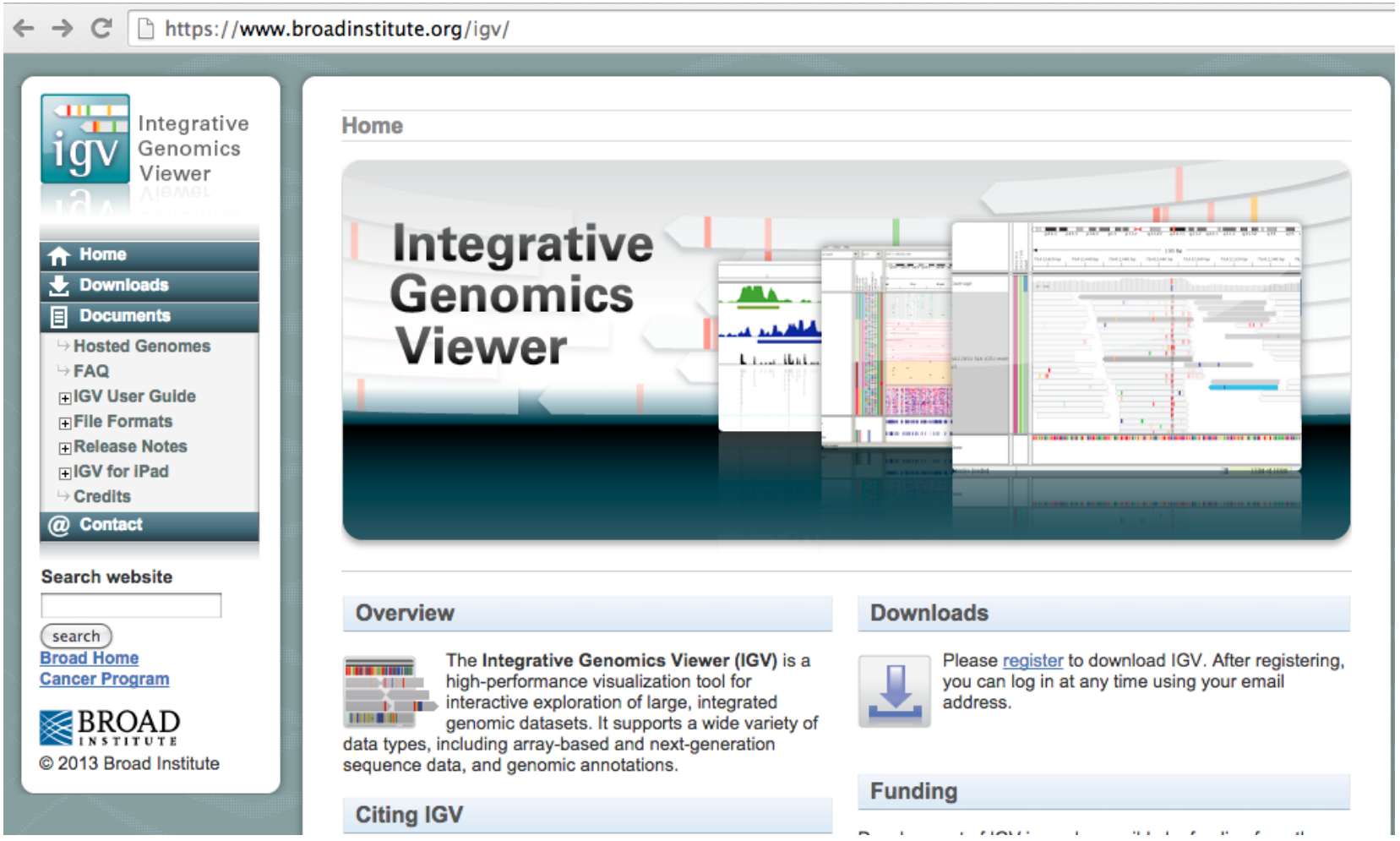
DMR Statistics



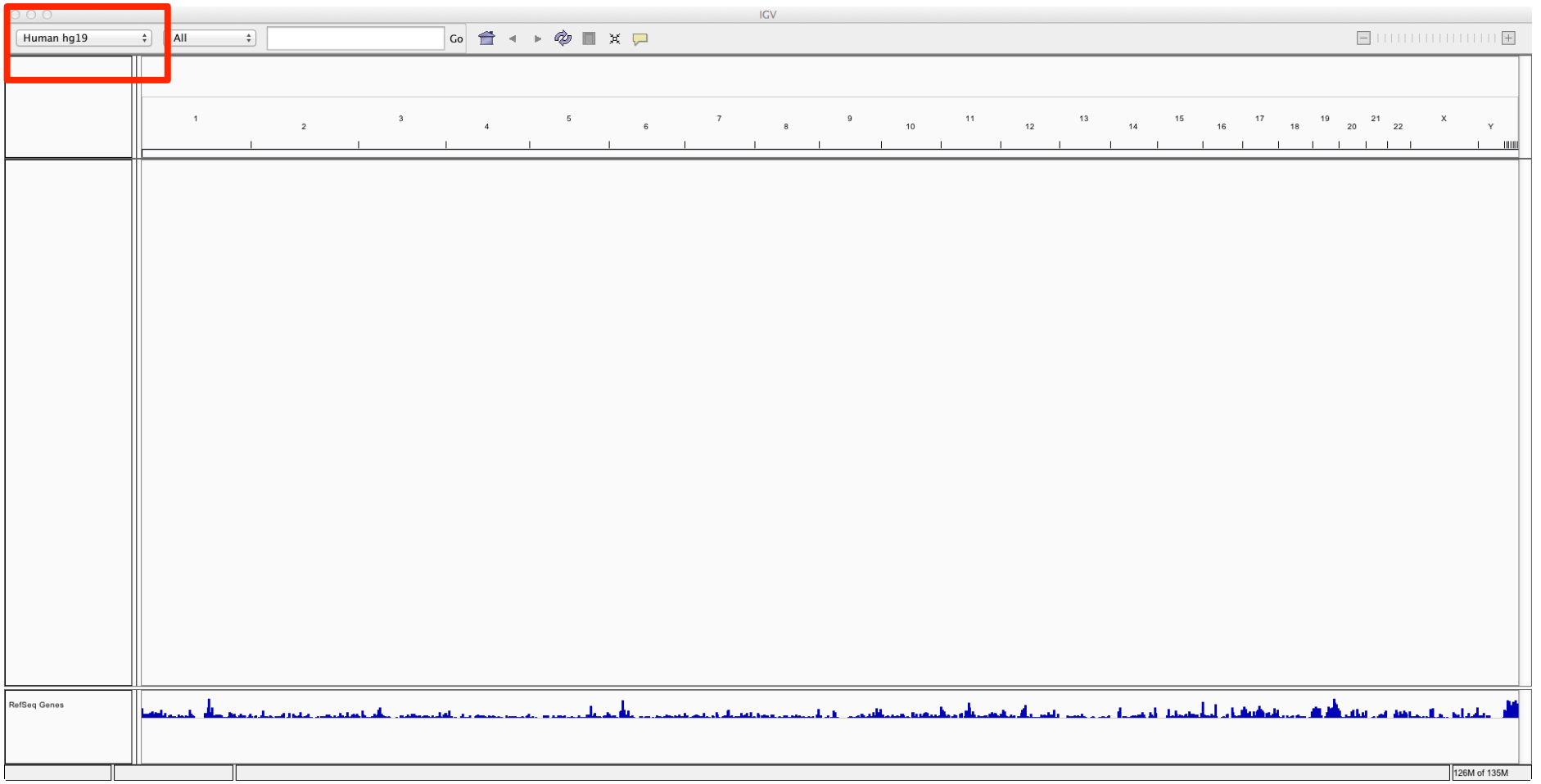


Visualization of DNA Methylation: IGV

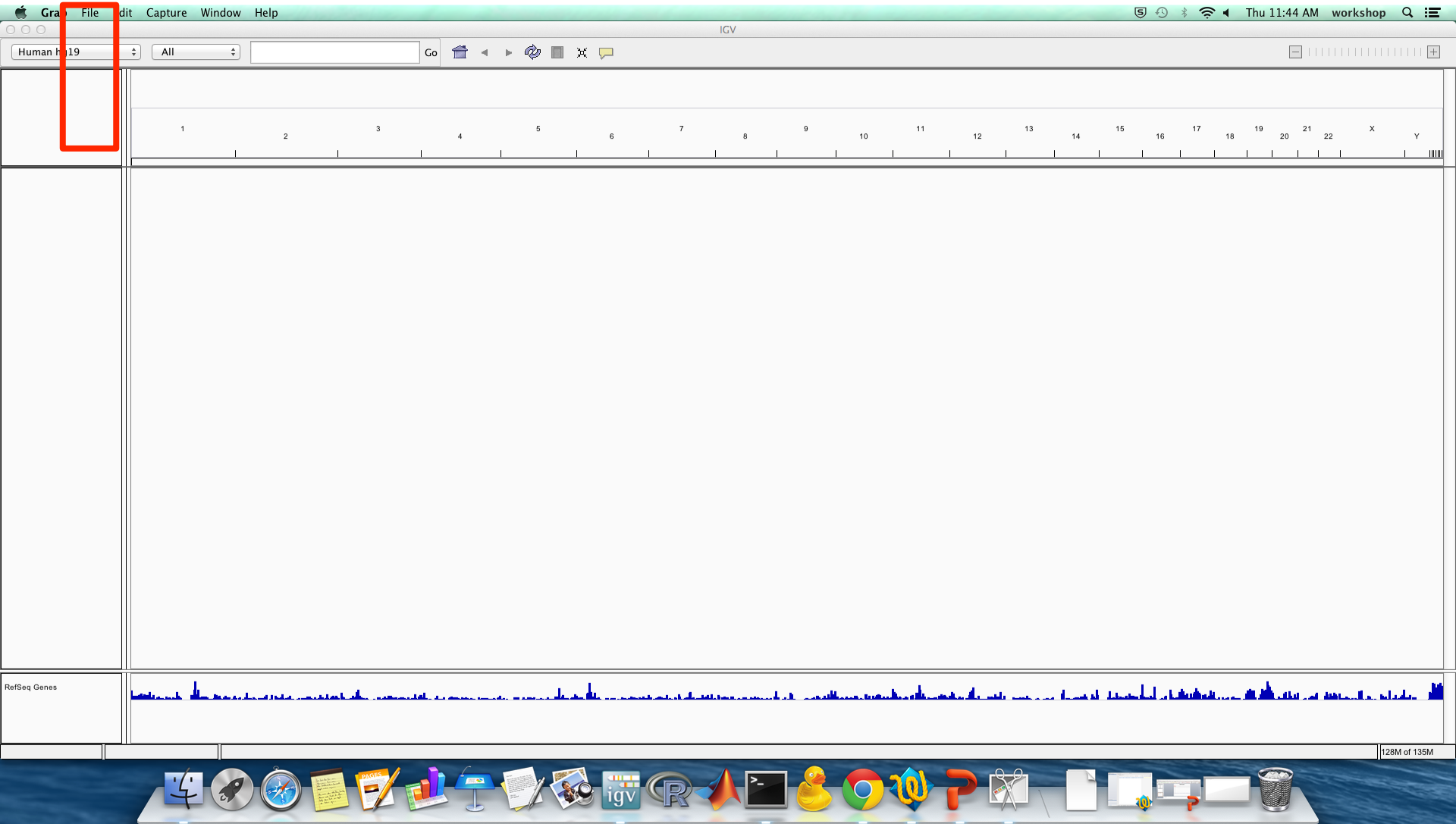
#download IGV Browser to your desktop
#IGV already contains built-in genomes for some common ones



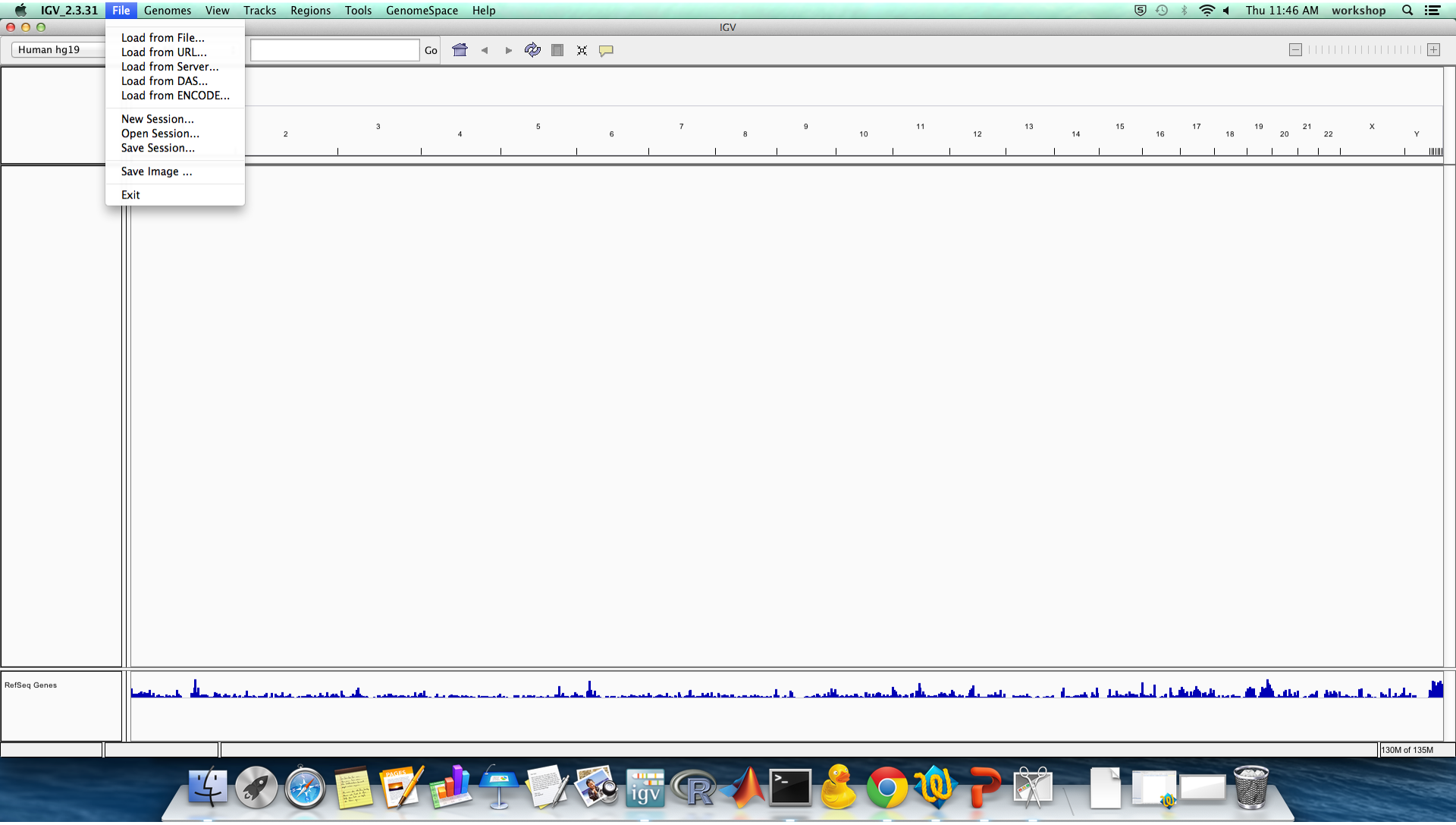
IGV Browser


IGV
Human hg19
All
Go
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
RefSeq Genes
126M of 135M

IGV Browser


File
Capture
Window
Help
Thu 11:44 AM
workshop
IGV
Human
All
Go
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
16
17
18
19
20
21
22
X
Y
RefSeq Genes
128M of 135M

IGV Browser


IGV_2.3.31
File
Genomes
View
Tracks
Regions
Tools
GenomeSpace
Help
Thu 11:46 AM
workshop
IGV
Human hg19
Go
Load from File...
Load from URL...
Load from Server...
Load from DAS...
Load from ENCODE...
New Session...
Open Session...
Save Session...
Save Image ...
Exit
RefSeq Genes
130M of 135M

Visualization File Formats

#For visualization on IGV, use bedgraph or bigwig files.
#convert .bed file output of bwa-meth/Bis-SNP to .bedgraph
#We did this already when preparing files for metilene

sed '1d' fileName.bed | awk '{print $1 "\t" $2 "\t" $3 "\t" $4}' > fileName.bedgraph

```
[flay@login4 raw]$ sed '1d' K562_bissnp.K562_R.meth.bed | awk '{print $1 "\t" $2 "\t" $3 "\t" $4}' > K562_bissnp.K562_R.meth.bedgraph
[flay@login4 raw]$ head K562_bissnp.K562_R.meth.bedgraph
chr1    847784  847784  0.0
chr1    1433528 1433528 100.0
chr1    1433539 1433539 100.0
chr1    1730072 1730072 100.0
chr1    1730163 1730163 100.0
chr1    1730204 1730204 100.0
chr1    1433574 1433574 100.0
chr1    1433638 1433638 100.0
chr1    1433677 1433677 100.0
chr1    3041726 3041726 0.0
```

Visualization File Formats

#Use UCSC Tools to convert between bedgraph to wig and bigwig formats.

http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/

#Tools are also available through Galaxy

UCSC Tools

#To install these tools, get the download link of the tool of interest:
#For example, in your installation directory:

wget http://hgdownload.soe.ucsc.edu/admin/exe/linux.x86_64/bedCoverage

#Make a specific file executable:

```
chmod +x bedCoverage
```

#or make all files in the directory executable:

```
chmod a+x
```

#Add installation directory to $PATH for easy access or specify full path/directory when calling the executable script

#For the tools we need today, they are already in:

```
/u/scratch/f/flay/workshop6/tools/ucsc/
```

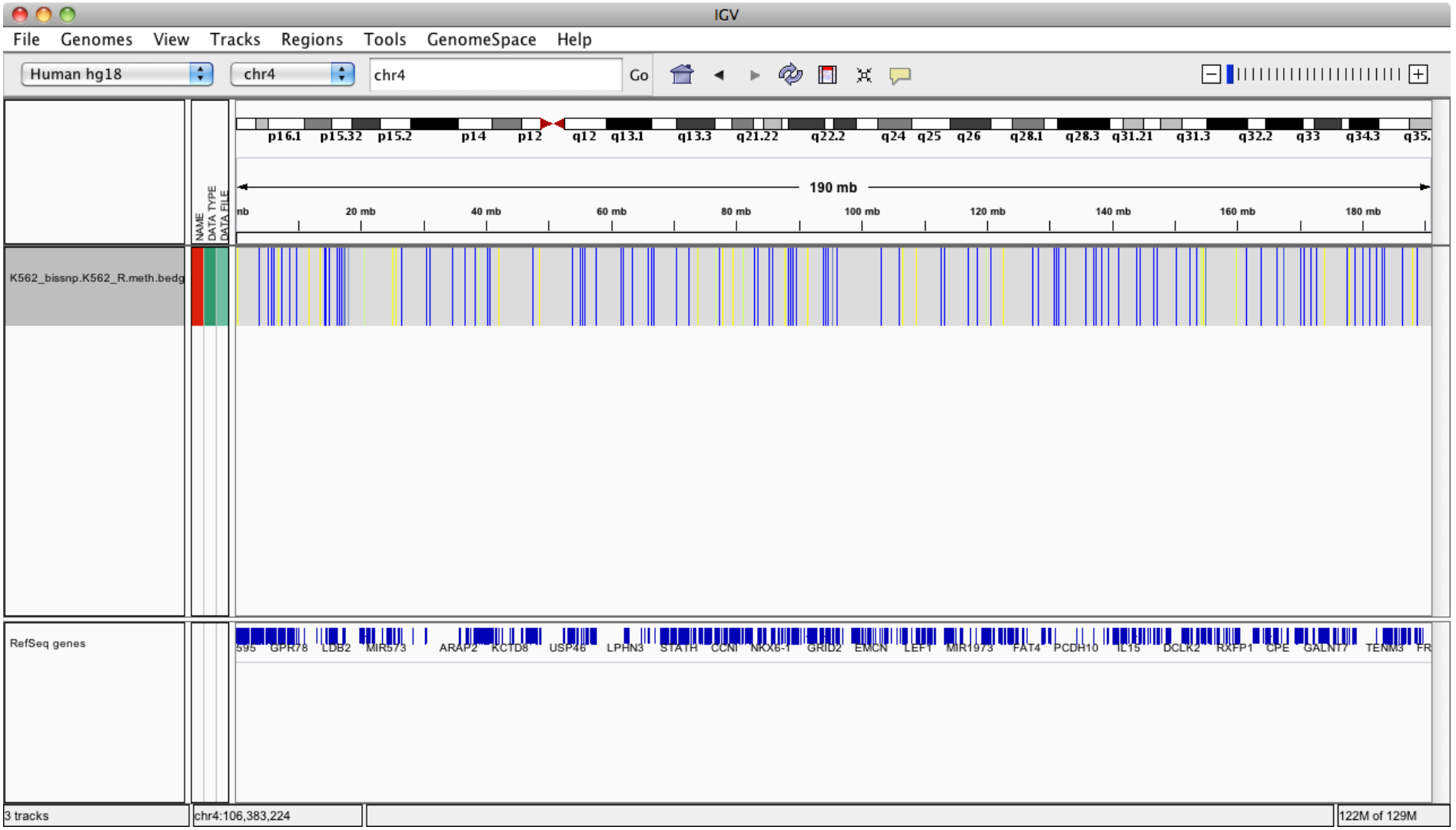
UCSC Tools:bedGraphToBigWig

Convert bedgraph to bigwig file
#Input bedgraph must be sorted

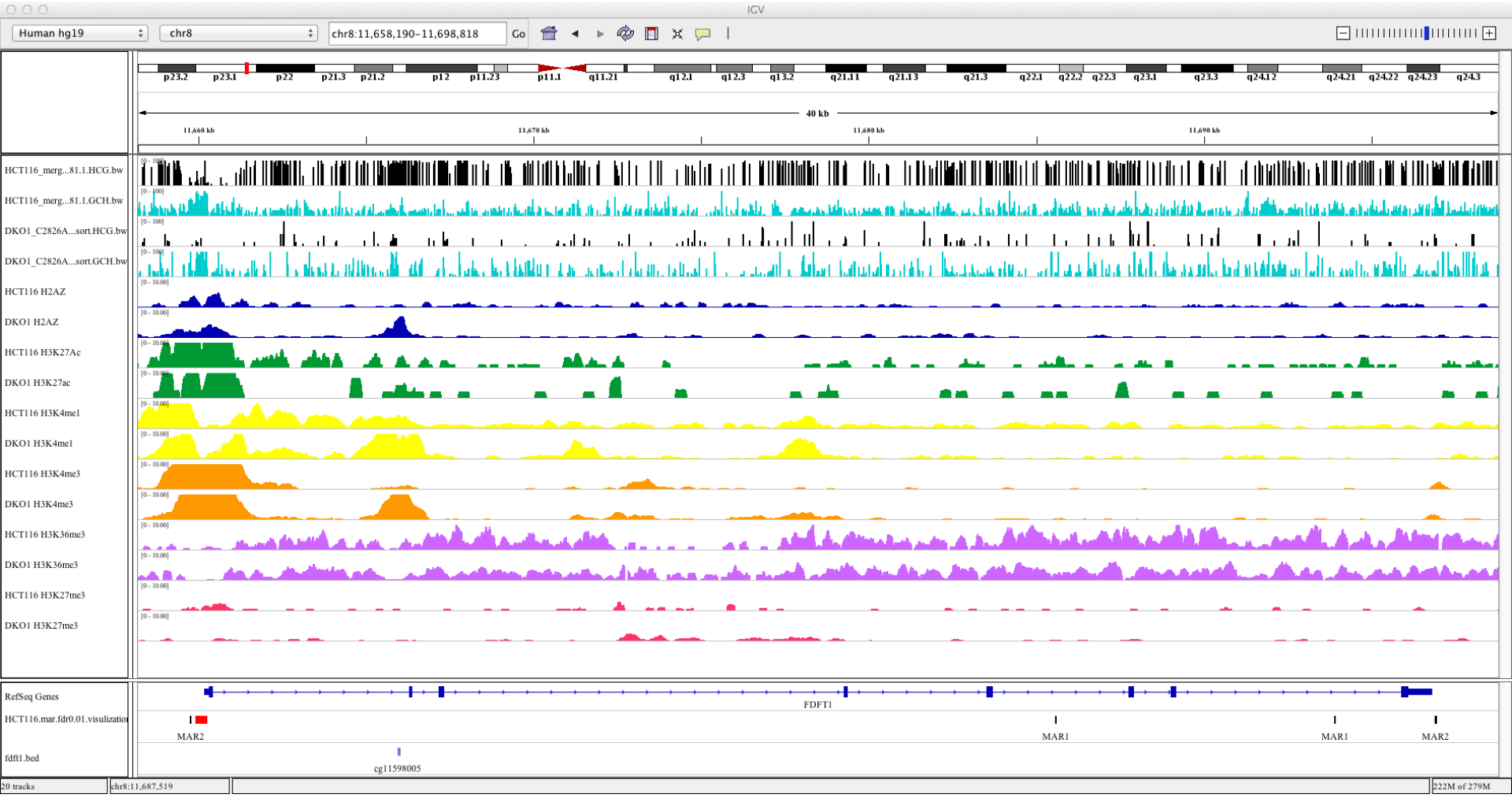
```
[flay@login4 raw]$ bedGraphToBigWig
bedGraphToBigWig v 4 - Convert a bedGraph file to bigWig format.
usage:
   bedGraphToBigWig in.bedGraph chrom.sizes out.bw
where in.bedGraph is a four column file in the format:
      <chrom> <start> <end> <value>
and chrom.sizes is two column: <chromosome name> <size in bases>
and out.bw is the output indexed big wig file.
Use the script: fetchChromSizes to obtain the actual chrom.sizes information
from UCSC, please do not make up a chrom sizes from your own information.
The input bedGraph file must be sorted, use the unix sort command:
  sort -k1,1 -k2,2n unsorted.bedGraph > sorted.bedGraph
options:
   -blockSize=N - Number of items to bundle in r-tree.  Default 256
   -itemsPerSlot=N - Number of data points bundled at lowest level. Default 1024
   -unc - If set, do not use compression.
```

bedGraphToBigWig fileName_sort.bedgraph /u/scratch/f/flay/workshop6/tools/ucsc/hg19.chrom.sizes fileName.bw

IGV Browser


IGV
File Genomes View Tracks Regions Tools GenomeSpace Help
Human hg18
chr4
chr4
Go
p16.1 p15.32 p15.2 p14 p12 q12 q13.1 q13.3 q21.22 q22.2 q24 q25 q26 q28.1 q28.3 q31.21 q31.3 q32.2 q33 q34.3 q35.
190 mb
nb 20 mb 40 mb 60 mb 80 mb 100 mb 120 mb 140 mb 160 mb 180 mb
NAME
DATA TYPE
DATA FILE
K562_bissnp.K562_R.meth.bedg
RefSeq genes
595 GPR78 LDB2 MIR573 ARAP2 KCTD8 USP46 LPHN3 STATH CCNI NKX6-1 GRID2 EMCN LEFT MIR1973 FAT4 PCDH10 IL15 DCLK2 RXFP1 CPE GALNT7 TENM3 FR
3 tracks
chr4:106,383,224
122M of 129M

IGV Browser


IGV
Human hg19
chr8
chr8:11,658,190-11,698,818
Go
40 kb
11,660 kb
11,670 kb
11,680 kb
11,690 kb
HCT116_merg...81.1.HCG.bw
HCT116_merg...81.1.GCH.bw
DKO1_C2826A...sort.HCG.bw
DKO1_C2826A...sort.GCH.bw
HCT116 H2AZ
DKO1 H2AZ
HCT116 H3K27Ac
DKO1 H3K27ac
HCT116 H3K4me1
DKO1 H3K4me1
HCT116 H3K4me3
DKO1 H3K4me3
HCT116 H3K36me3
DKO1 H3K36me3
HCT116 H3K27me3
DKO1 H3K27me3
RefSeq Genes
FDFT1
HCT116.mar.fdr0.01.visulizatio
MAR2
MAR1
MAR1
MAR2
fdft1.bed
cg11598005
20 tracks
chr8:11,687,519
222M of 279M

Make Heatmap

```
#Use R to make heatmap – R is covered in Workshop 3
    module load R

#call R by typing R in command line
#load R packages when R is booted:
    library('RColorBrewer')
    library('gplots')

#to read the methylation data into R
met = read.delim("metilene_BL_FL.input", sep="\t")

#make a data matrix file
met.ma = as.matrix(met[,3:20])
```

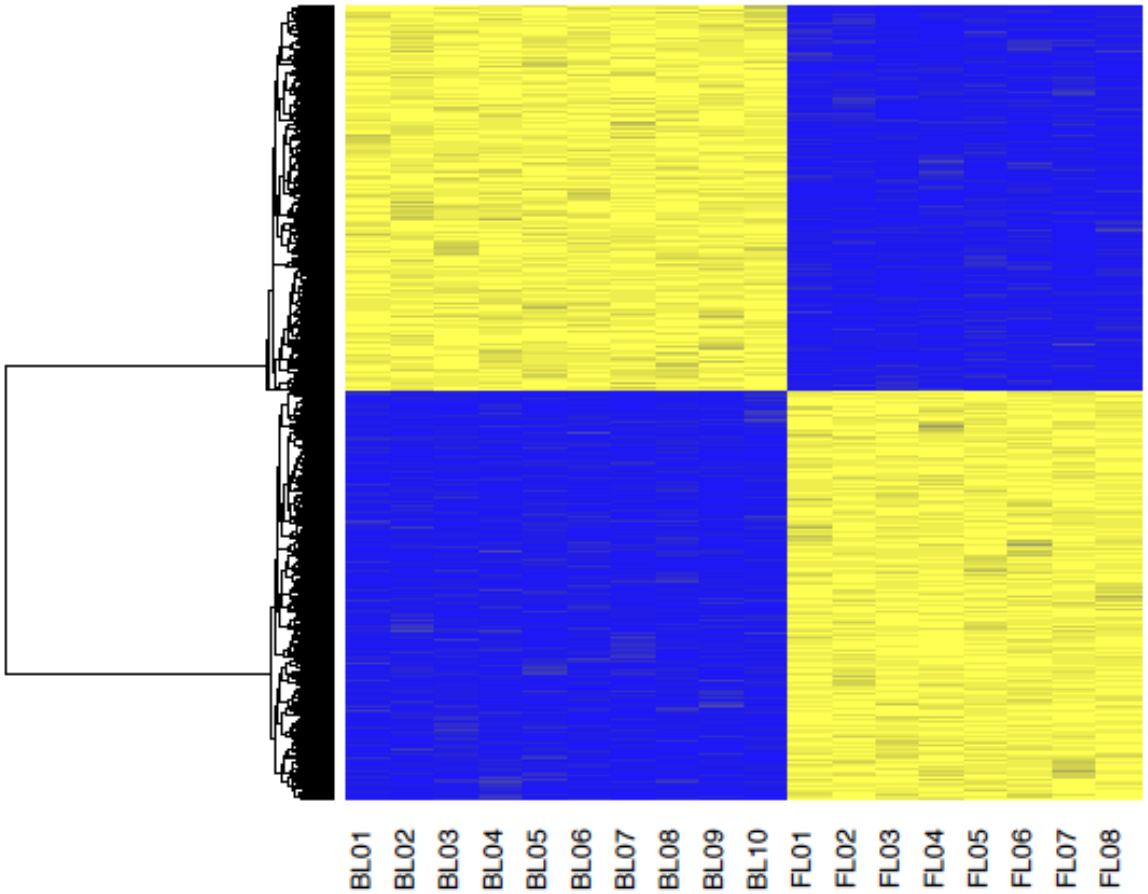
Make Heatmap

```
#calculate variance across all samples in each CG site.
met.ma.var = apply(met.ma,1,var)

#extract the top 1% most variable CG sites in the dataset and write
as a new matrix
met.var = met.ma[met.ma.var>quantile(met.ma.var,0.99),]

#perform clustering of data matrix
met.hclust=hclust(dist(met.var))

#Generate a heatmap

pdf(“heatmap.pdf”)
heatmap.2(met.var,col=colorRampPalette(c("blue","yellow")), Colv =
"FALSE", dendrogram= "row", Rowv=as.dendrogram(met.hclust))
dev.off()
```

Make Heatmap

#download pdf on to your desktop

scp -r flay@hoffman2.idre.ucla.edu:/u/scratch/f/flay/workshop6/data/raw/heatmap.pdf ~/Desktop



Make Heatmap

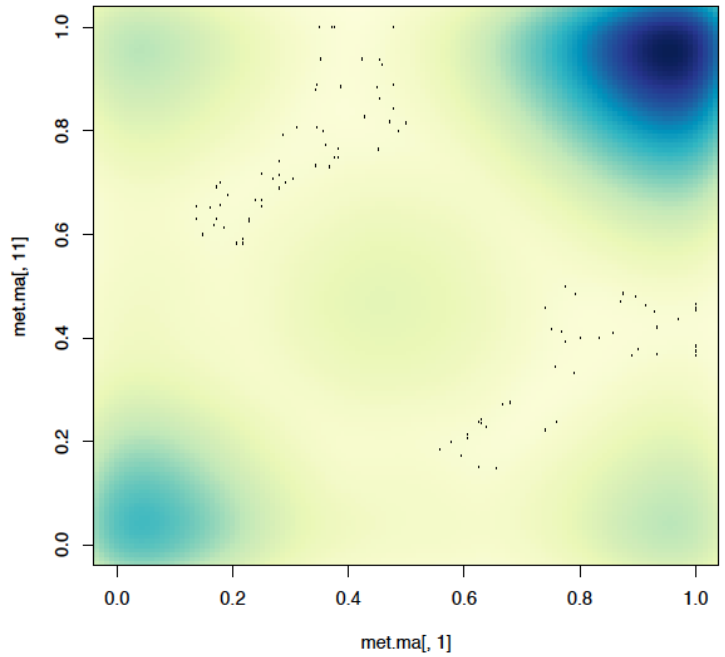
#To alter plotting parameters, check the various arguments for the function by typing:

```
?heatmap.2
```

Smooth Scatter Density Plot

#Use a built-in function smoothScatter

```
pdf("smoothScatter.pdf")
smoothScatter(met.ma[,1], met.ma[,11],
colramp=colorRampPalette(brewer.pal(8,"YlGnBu")))
dev.off()
```



Colors in R Plot

#brewer.pal is a function of RColorBrewer package
#For more brewer.pal color options:
http://www.personal.psu.edu/cab38/ColorBrewer/ColorBrewer.html

#Other color options:
https://www.nceas.ucsb.edu/~frazier/RSpatialGuides/colorPaletteCheatsheet.pdf

Other Commonly Used WGBS Tools

Aligners and Methylation Callers:

- BSMAP v2.9 https://code.google.com/archive/p/bsmap/
- BS-Seeker2 http://pellegrini.mcdb.ucla.edu/BS_Seeker2/
- Bis-mark http://www.bioinformatics.babraham.ac.uk/projects/bismark/

DMR Callers:

- bsseq http://bioconductor.org/packages/release/bioc/html/bsseq.html
- MOABS https://code.google.com/archive/p/moabs/