Visualize your data

IGV: Visualize methylation data in IGV genome browser

- You can download IGV genome browser here:

http://www.broadinstitute.org/software/igv/download

- Hosted genomes:

http://www.broadinstitute.org/software/igv/Genomes

- Add your own tracks. Can add multiple tracks and compare:

  - Genomes -> Load genome from server

  - Genomes -> Load genome from file

IGV: Visualize methylation data in IGV genome browser

- Visualize BS-Seeker output files:

    - Methylation levels in wiggle file (ex: **file1.wig**)

    - Aligned reads in .bam_sorted.bam file (ex: **file1.bam_sorted.bam** and **file1.bam_sorted.bam.bai**, both must be in same directory)
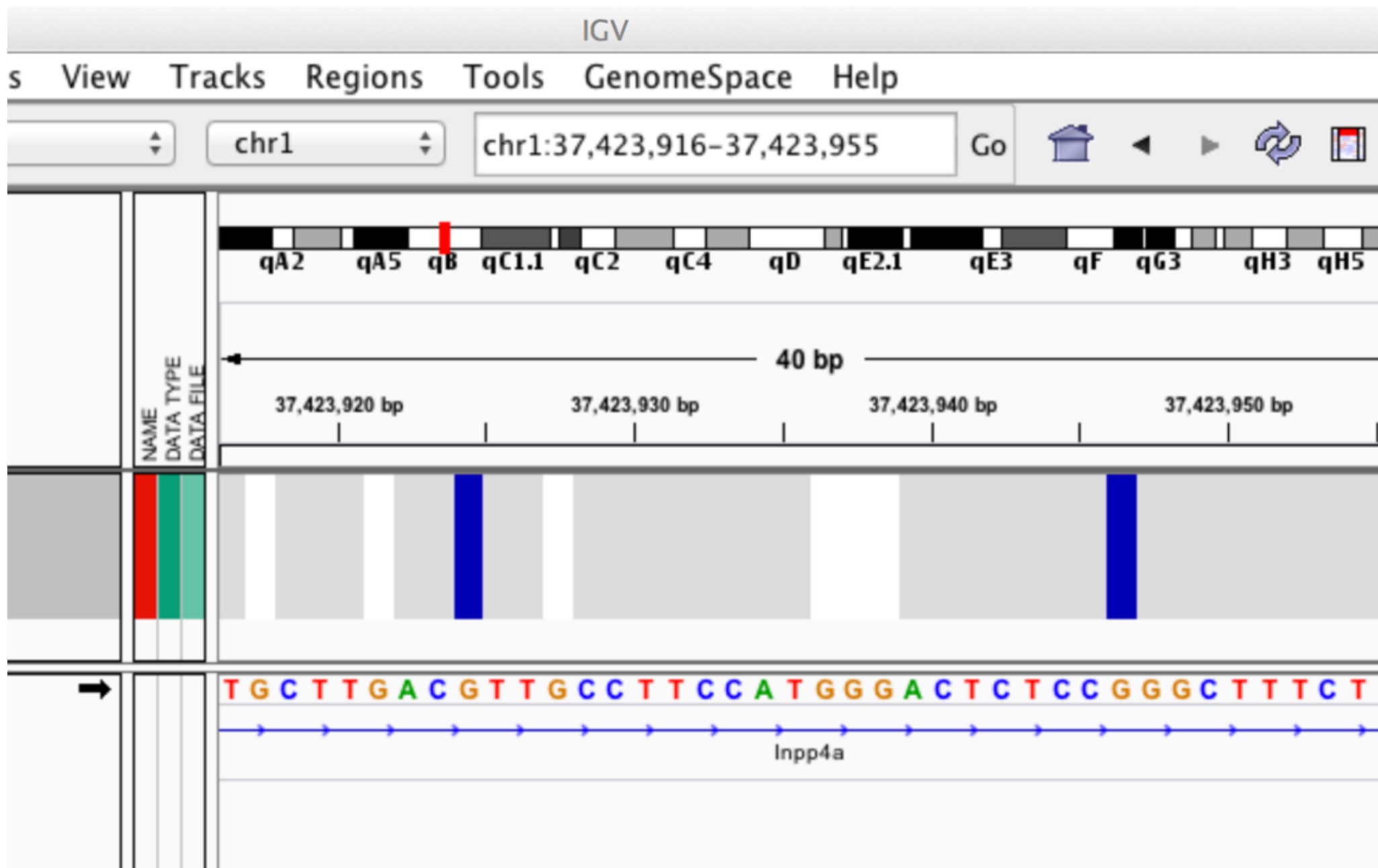
    - File -> Load from File

IGV: wiggle file **file1.wig**
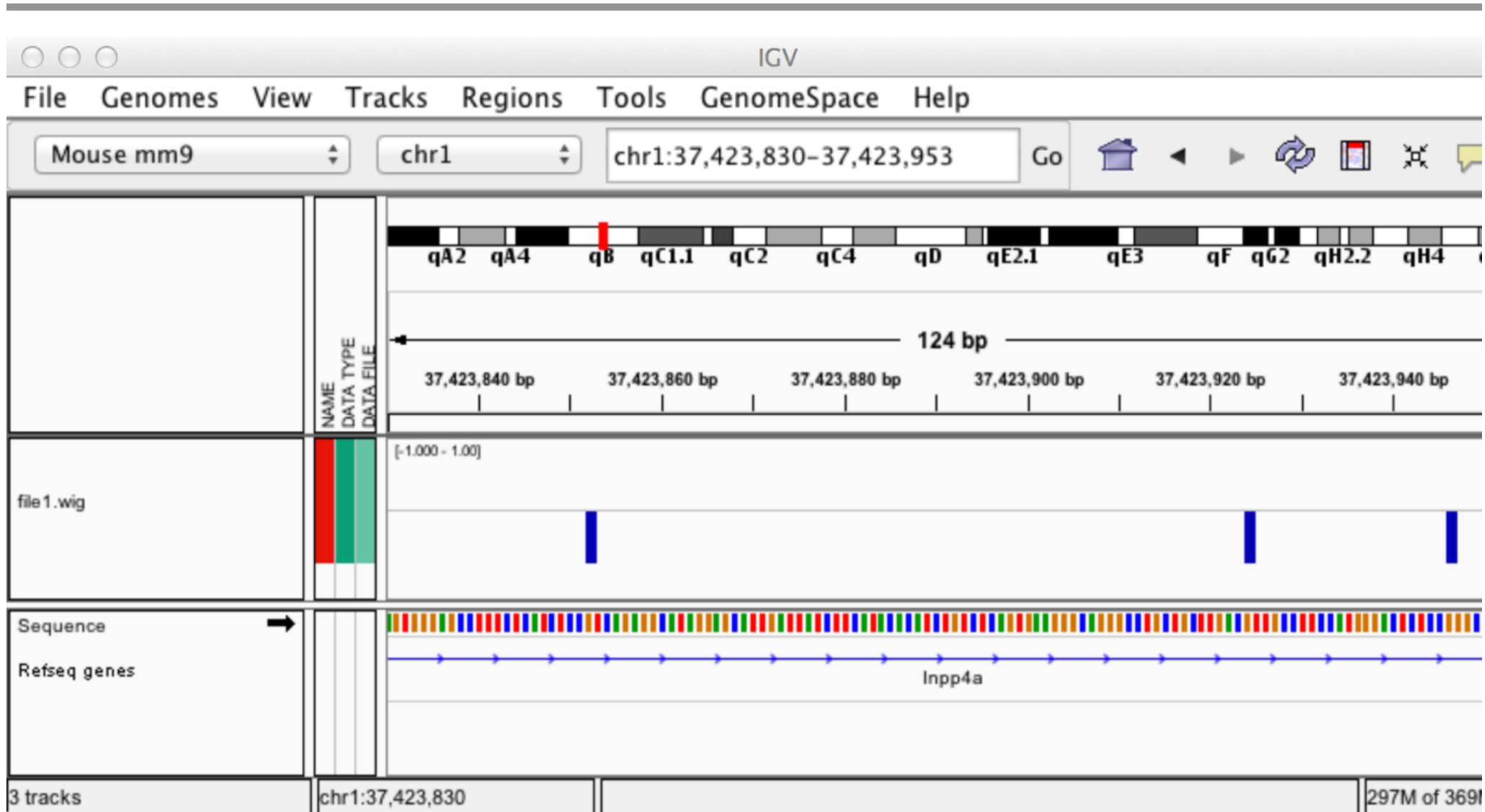
IGV: Visualize methylation data in IGV genome browser

- Change Graph

  - Right click on track data

  - Select: **Heatmap**, Bar chart, Points, Line plot

- In Heatmap, **missing** data is represented in grey, **zero** values are in white
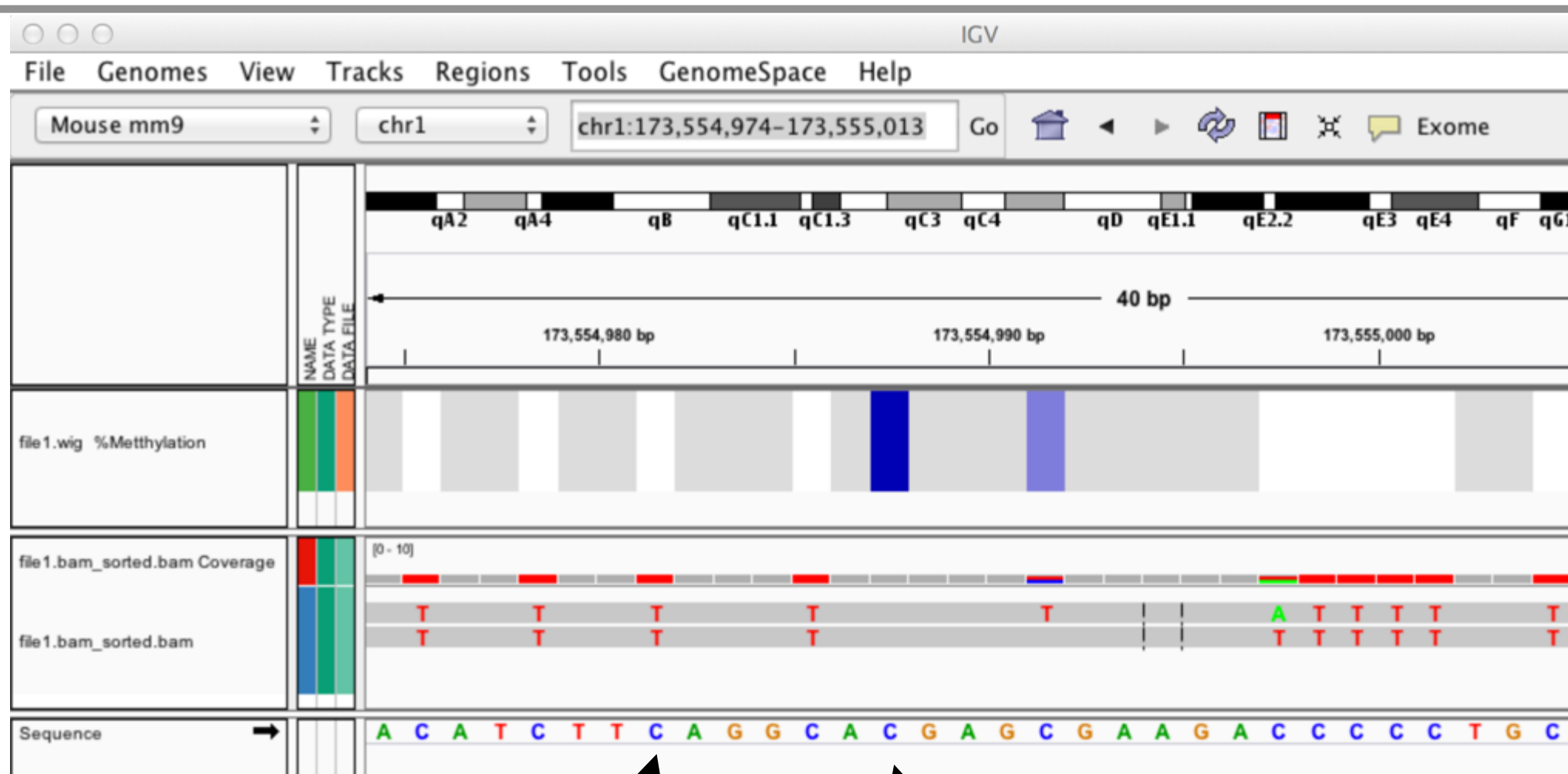
IGV: Visualize methylation data in IGV genome browser

IGV: Load bam file **file1.bam_sorted.bam**

IGV: Visualize methylation data in IGV genome browser

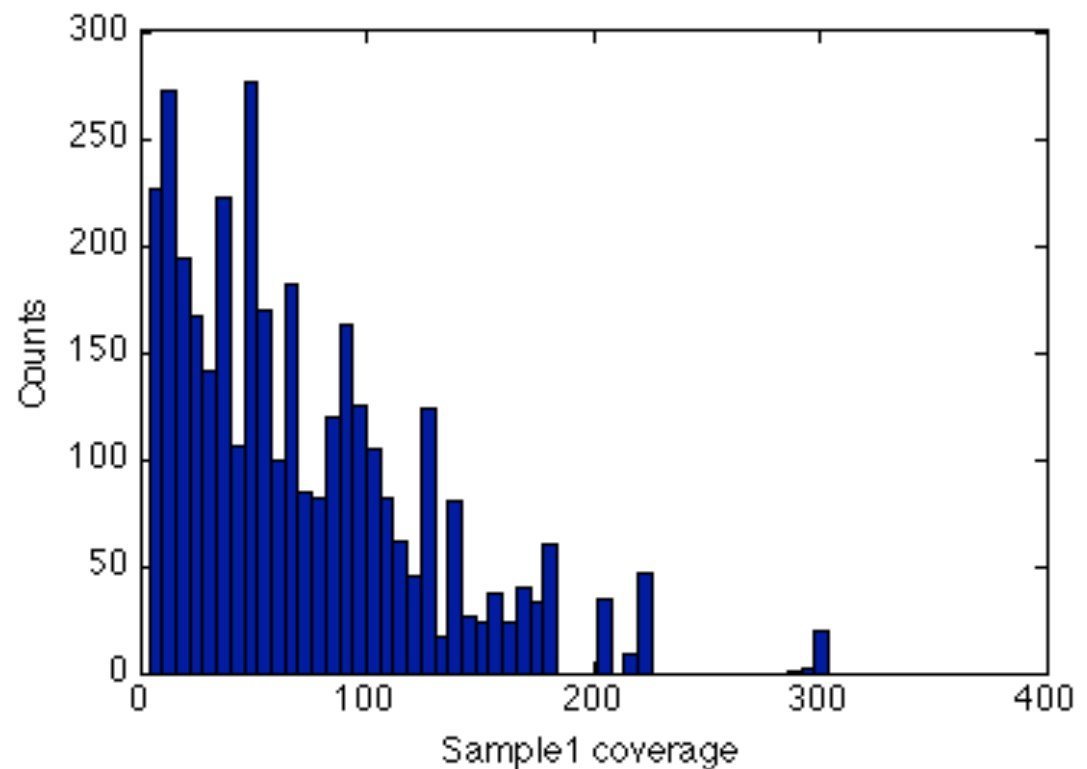Reference: C
Read: T
Not methylated

Reference: C
Read: C
Methylated

Plot data in MATLAB

Plots: Histogram
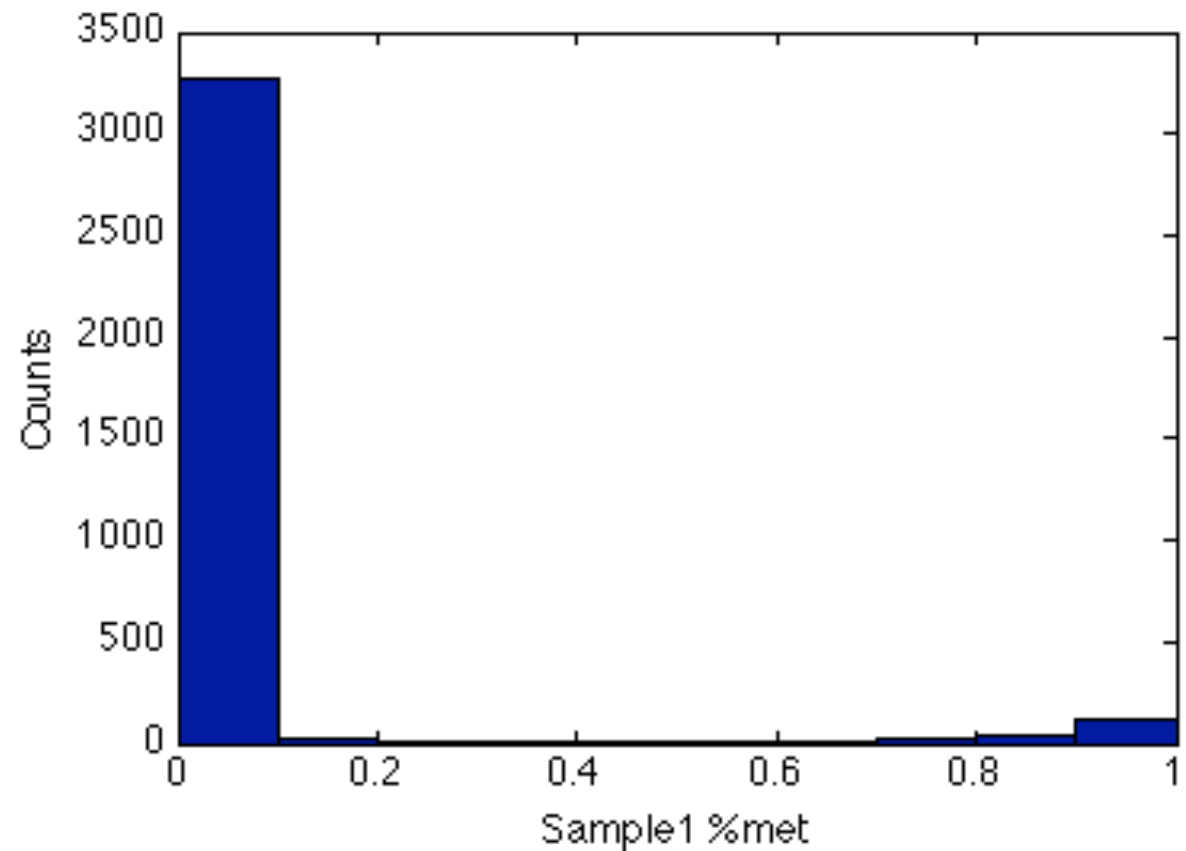
```
%Coverage (total counts)
x = sum(Sample1_data(:,4:5),2);

%Histogram of coverage
figure; hist(x)
figure; hist(x,50)
```

Plots: Histogram

```
% Methylation level(%)
y = Sample1_data(:,3);

% Histogram of %methylation
figure; hist(y)
```

Plots: Histogram
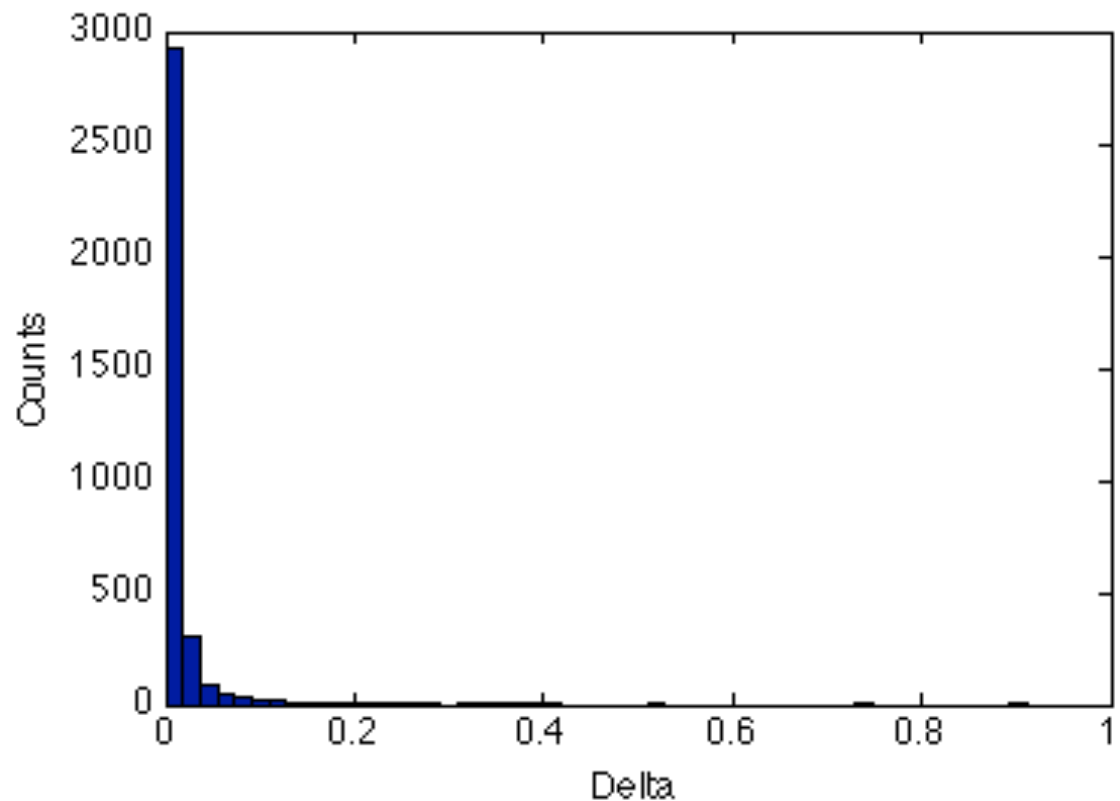
```
%Delta for Sample1 and Sample2
p1 = Sample1_data(:,3);
p2 = Sample2_data(:,3);
delta = abs(p1-p2);

% Histogram of delta
figure; hist(delta)
```
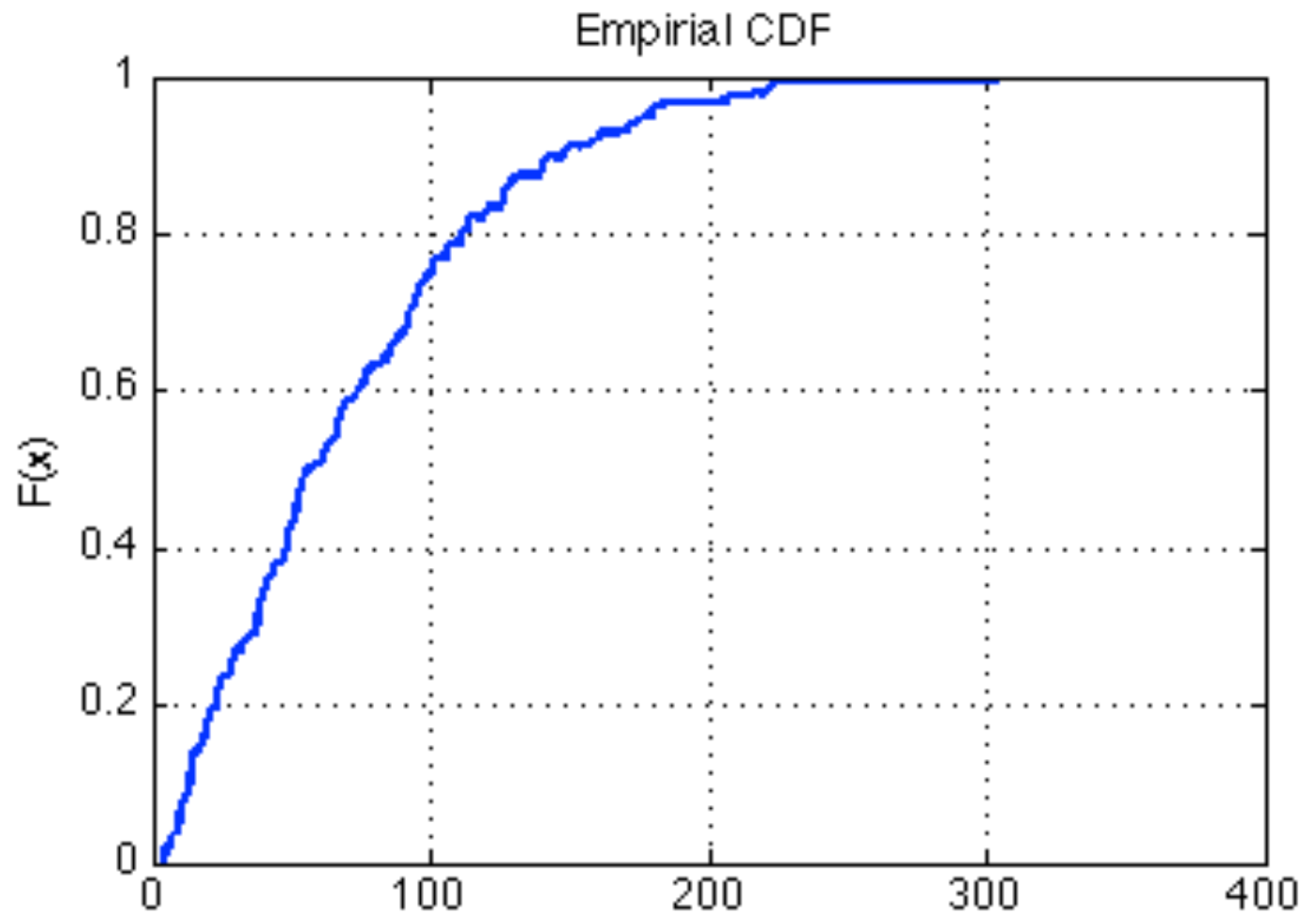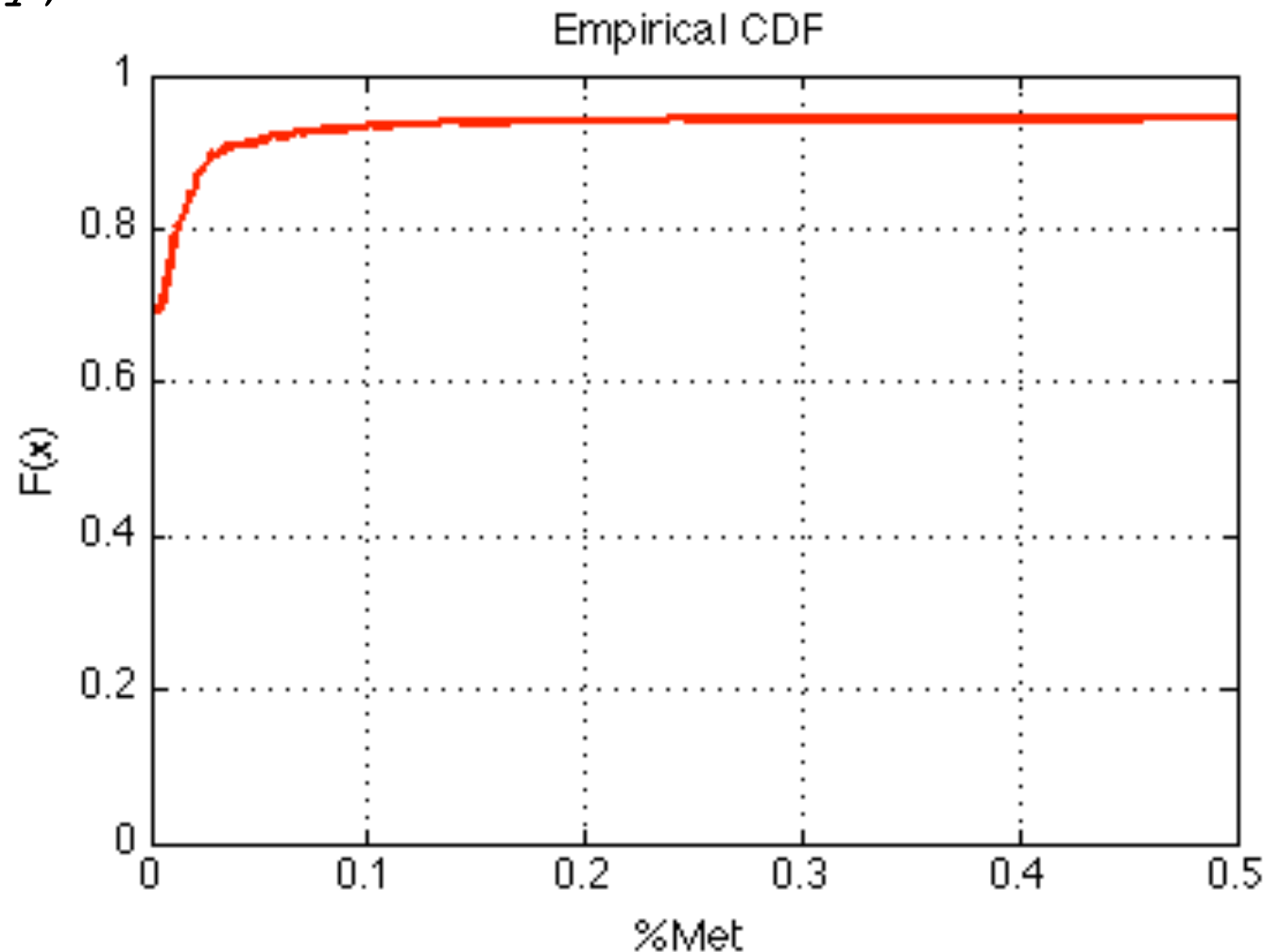
Plots: Commutative distribution function (CDF) plot

```
%CDF plot of coverage
figure; cdfplot(x)
```


Empirial CDF

Plots: Commutative distribution function (CDF) plot
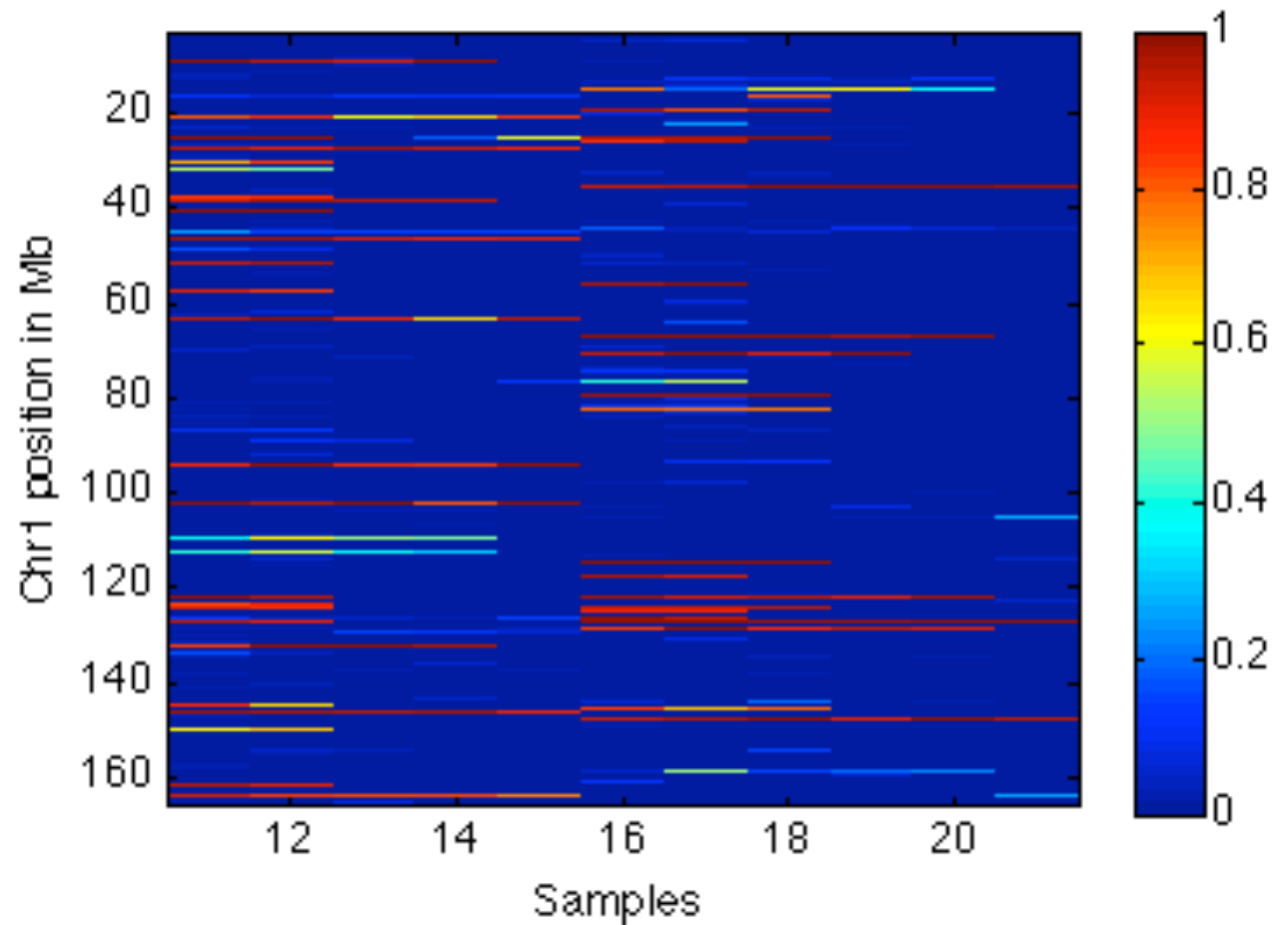
```
% CDF plot of %methylation
figure; cdfplot(y)
```

Plots: Heat map (imagesc)

```
%Select %met data
data = [ConditionA_data(:,3:end),ConditionB_data(:,3:end)];
```

```
%Heatmap of data
imagesc(data)
```



```
imagesc(11,ConditionA_data(:,2)./1e06, data)
```

Plots: Clustergram

```matlab
   Select %met data
 data = [ConditionA_data(:,3:end),ConditionB_data(:,
 3:end)];

 %Labels for data
 labels = {'A','A','A','A','A',...
           'B','B','B','B','B','B',};
 labels = [repmat({'A'},1,5), repmat({'B'},1,6)];

 %Index for C-sites with no missing values
 a = sum(isnan(data),2)==0;

 %Clustergram of data
 clustergram(data(a,:),'ColumnLabels', labels)

 %Clustergram of data
 clustergram(data(a,:), 'ColumnLabels',
 labels,'Symmetric', false)
```

Plots: Clustergram

```matlab
   Select %met data
data = [ConditionA_data(:,3:end),ConditionB_data(:,
3:end)];

%Labels for data
labels = {'A','A','A','A','A',...
          'B','B','B','B','B','B',};
labels = [repmat({'A'},1,5), repmat({'B'},1,6)];

%Index for C-sites with no missing values
a = sum(isnan(data),2)==0;

%Clustergram of data
clustergram(data(a,:),'ColumnLabels', labels)
```
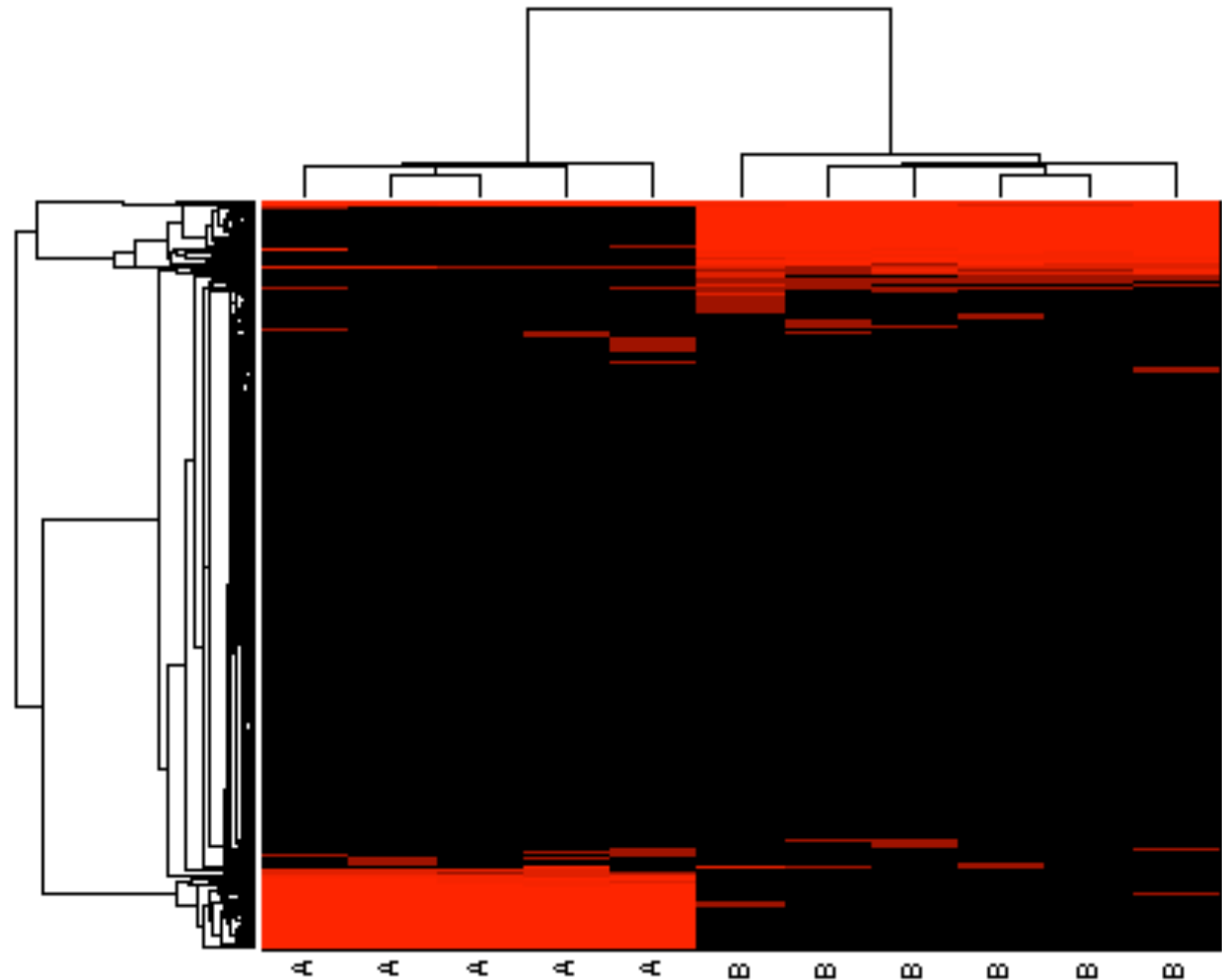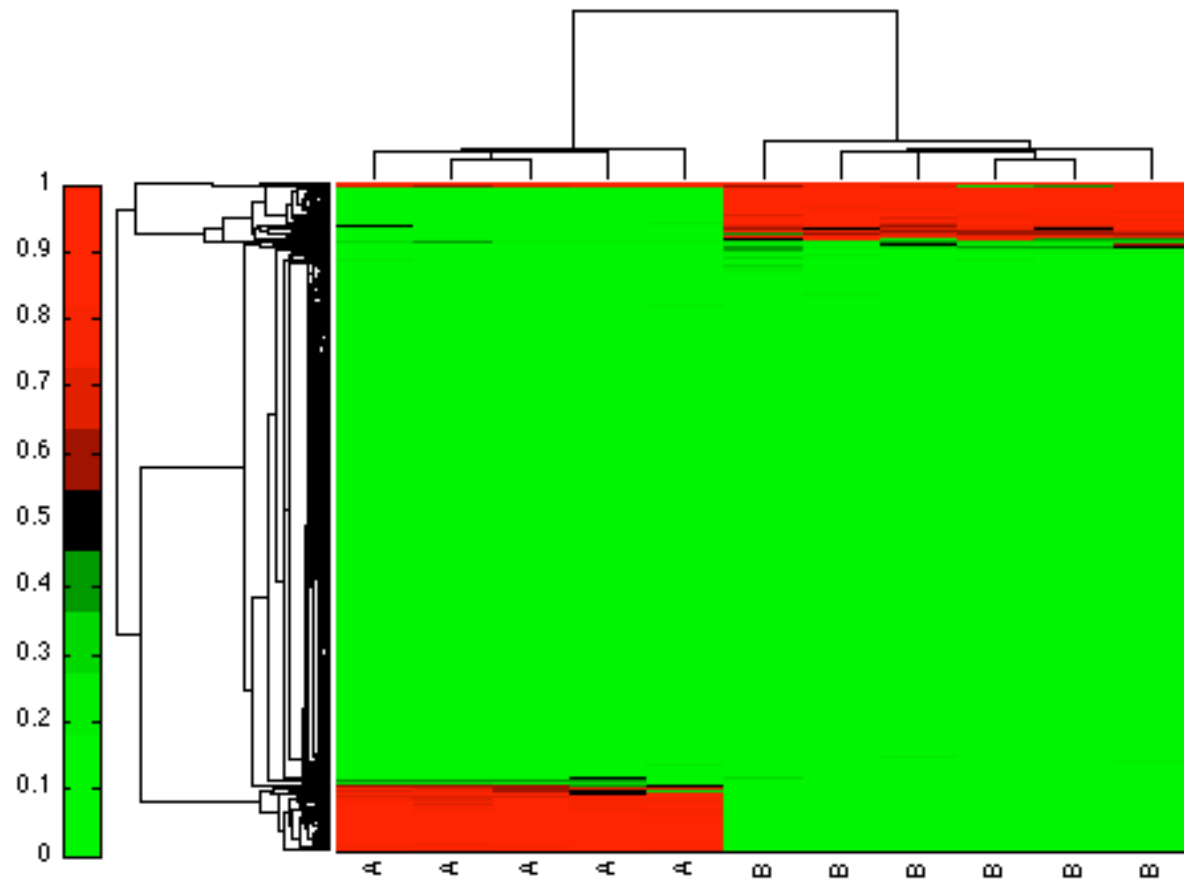
Plots: Clustergram

```
%Clustergram of data
clustergram(data(a,:),'ColumnLabels', labels)
```

Plots: Clustergram

```
%Clustergram of data
clustergram(data(a,:), 'ColumnLabels',
labels,'Symmetric', false)
```

Plots: Anova and box plot

```matlab
  %Data and grouping variable
data;
labels = [repmat({'A'},1,5), repmat({'B'},1,6)];

%Allocate matrix for output p-val
p_anova = zeros(length(data),1);

%Anova test
for i=1:length(data)
    p_anova(i) = anova1(data(i,:),labels, 'off');
end

%Minimum p-value
[m,in] = min(p_anova)

%Plot for that site
anova1(data(in,:),labels)
```
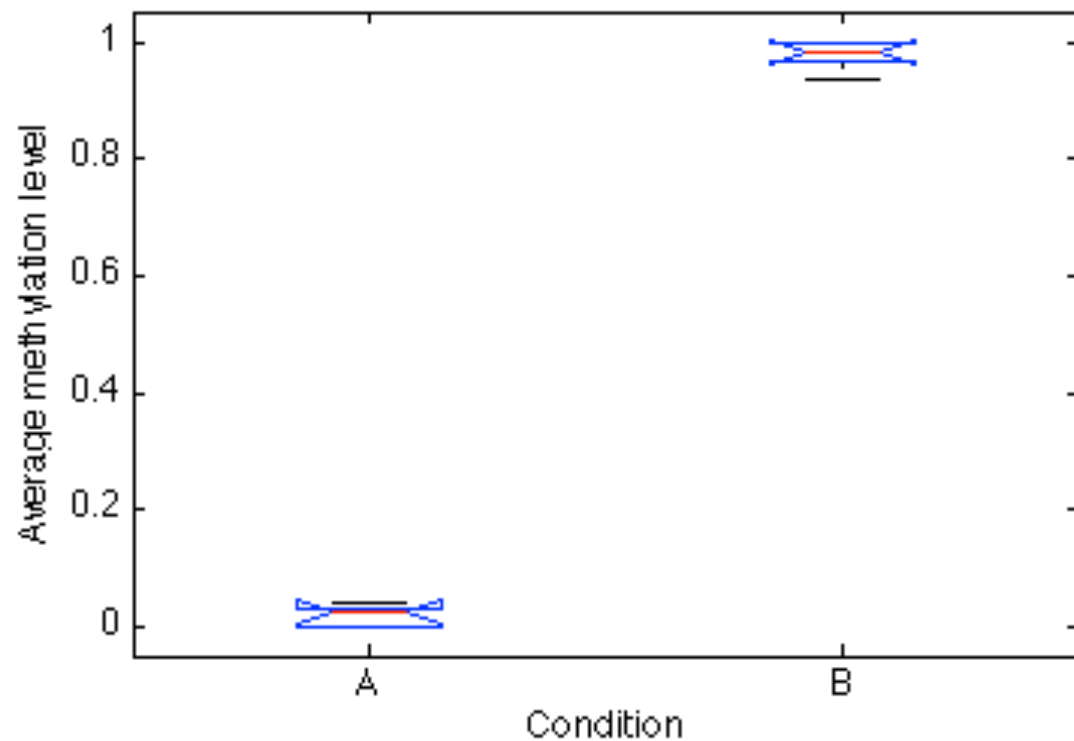
```
%Plot for that site
anova1(data(in,:),labels)
```
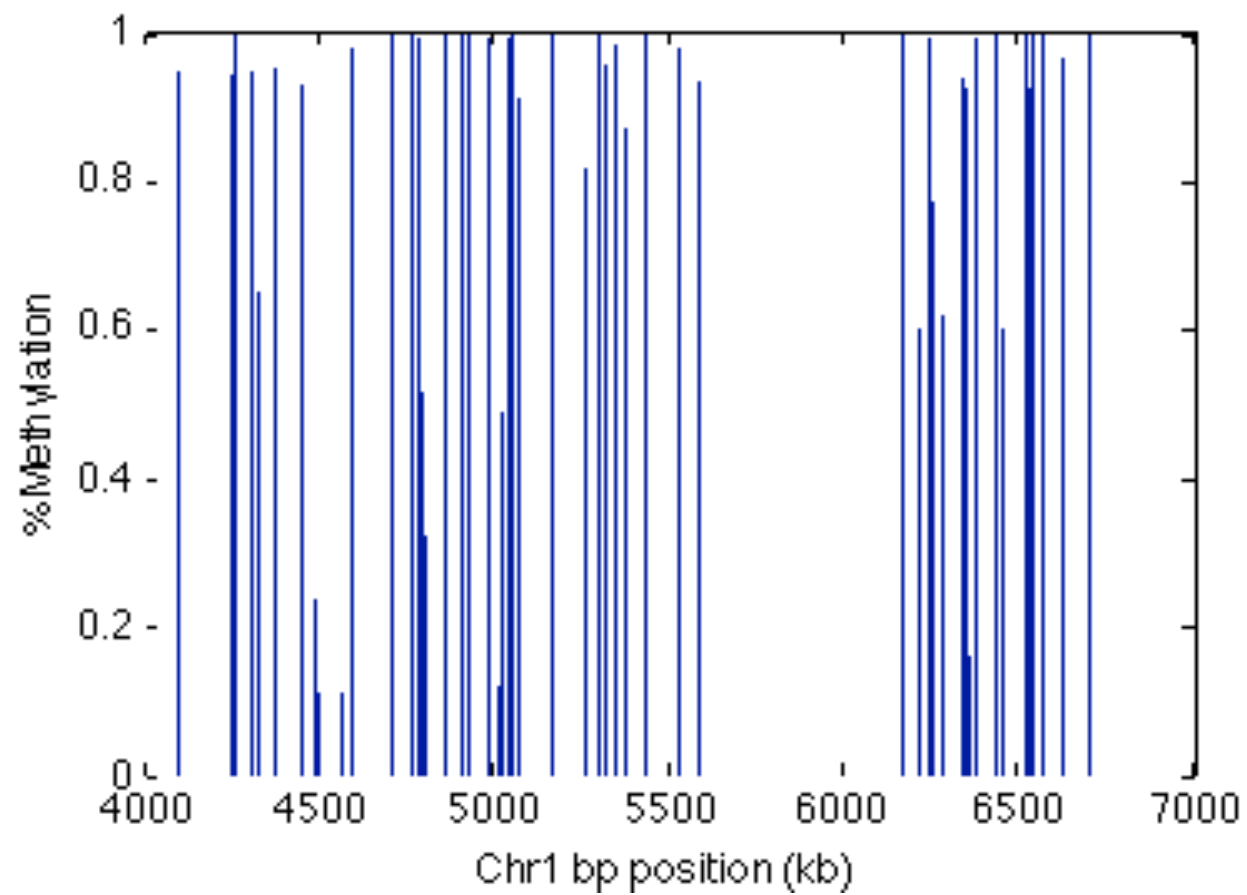
Plots: Bar plot

```matlab
   %Data for x and y axes
x = Sample1_data(:,2)./1e03;     % Bp position in kb
y = Sample1_data(:,3);           % Met level

%Bar plot
figure;
bar(x,y)

%Labels for plot
xlabel('Chr1 bp position (kb)')
ylabel ('%Methylation')
```

Plots: Bar plot

`bar(x,y)`

Plots: Line plot & Bar plot for two samples

```
%Plot methylation level for C-site with
minimun p-val in t-test
[m,in] = min(p);

%x-axis
x2 = [1,2];

%y-axis is mean %Met for each condition
y2 = [nanmean(ConditionA_data(in,3:end)),
nanmean(ConditionB_data(in,3:end))];
```
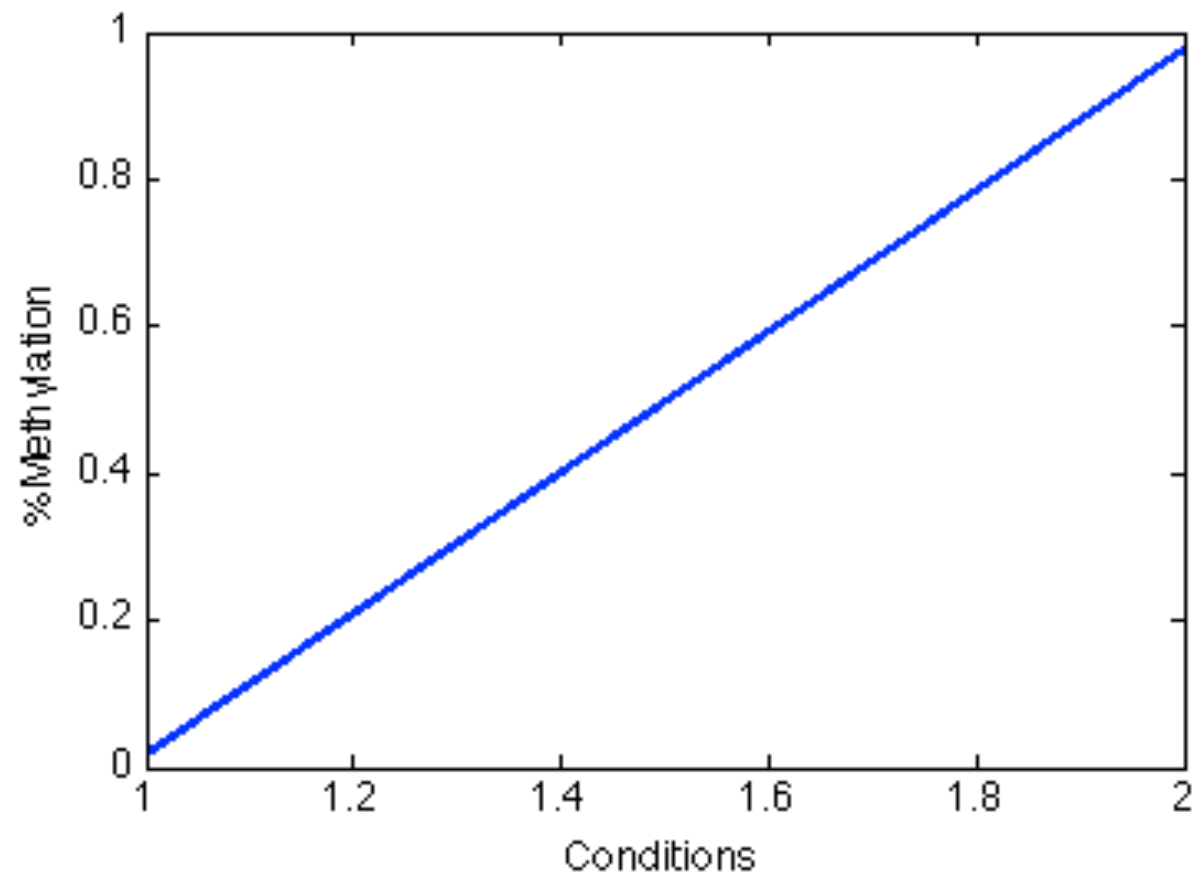
Plots: Line plot & Bar plot for two samples

```
%Line plot
plot(x2,y2)
```
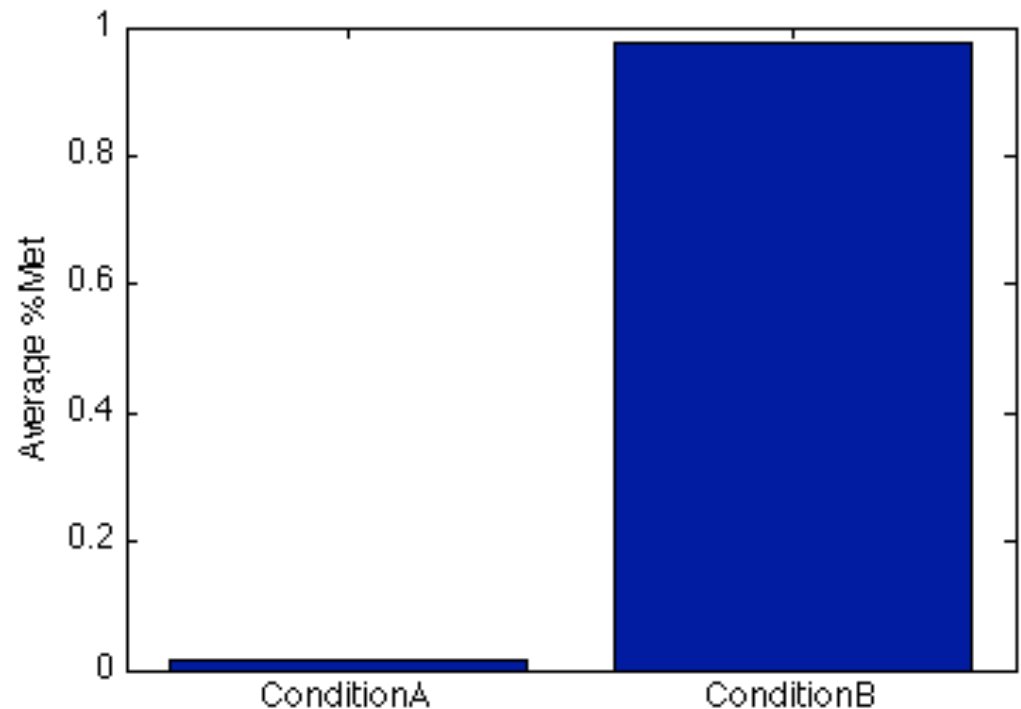
Plots: Line plot & Bar plot for two samples

```matlab
%Bar plot
figure;
bar(x2,y2)

set(gca, 'XTickLabel', {'ConditionA',
'ConditionB'})
ylabel('Average %Met')
```

Plots: Scatter plot

```matlab
%Data for x and y axes
x = Sample1_data(:,2)./1e03;    % Bp position in kb
y = Sample1_data(:,3);          % Met level

%Scatter plot
scatter(x,y)

scatter(x,y,'k','filled')

scatter(x,y, 'm', '*')

scatter(x,y, 'r', 'd', 'filled')
```
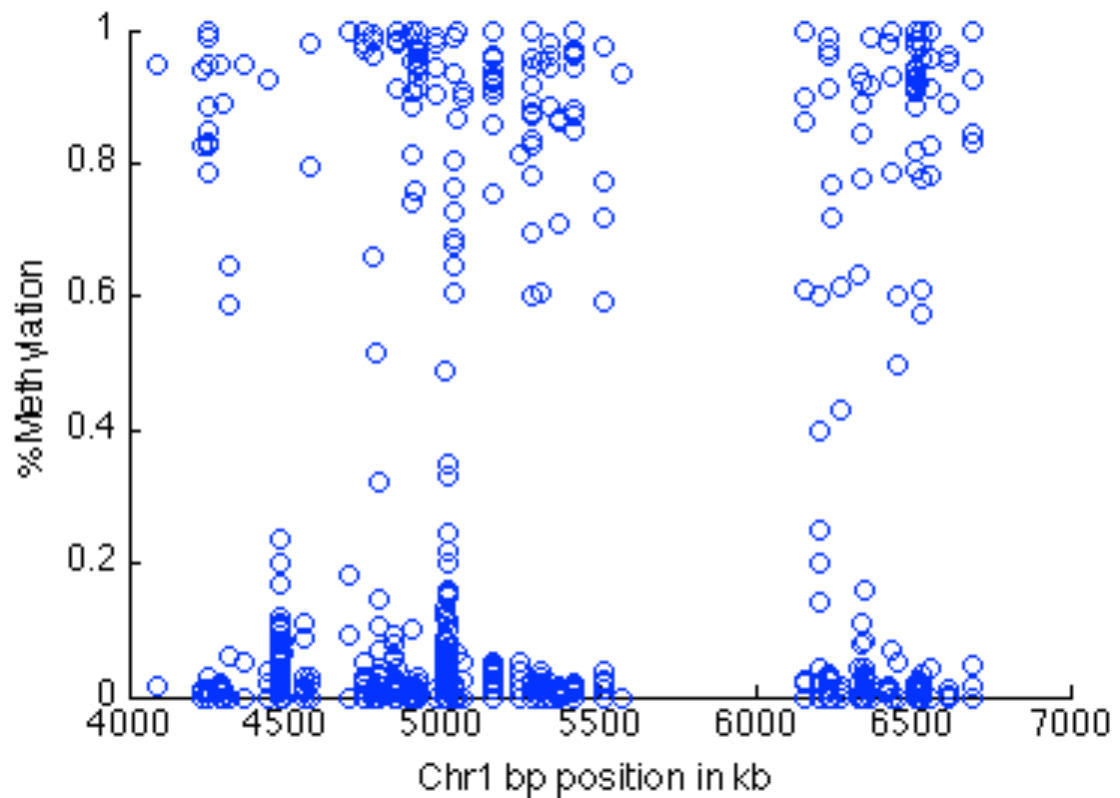
Plots: Scatter plot

```
%Scatter plot
scatter(x,y)
xlabel('Chr1 bp position in kb')
ylabel('%Methylation')
```
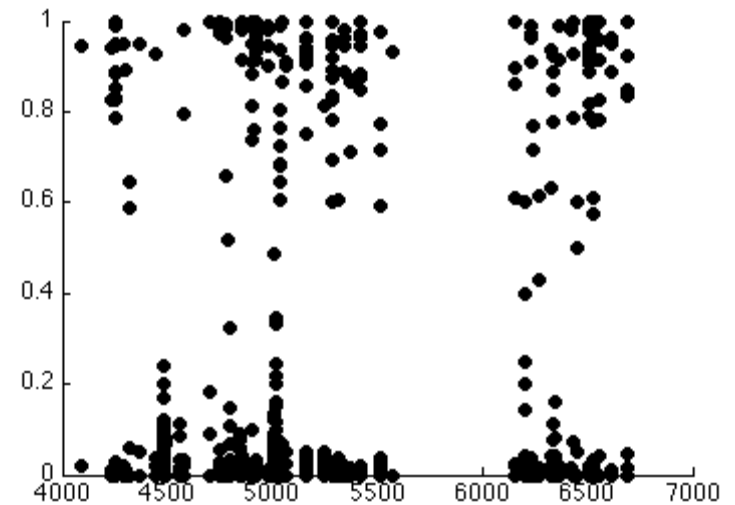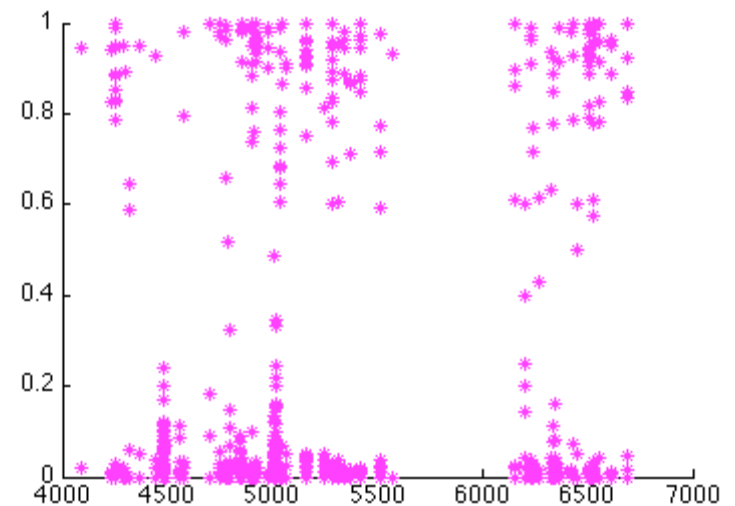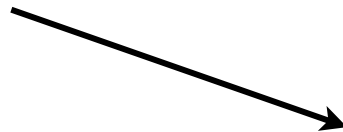
Plots: Scatter plot

```
%Scatter plot
scatter(x,y)
xlabel('Chr1 bp position in kb')
ylabel('%Methylation')
```

Plots: Scatter plot

scatter(x,y,'k','filled')

scatter(x,y, 'm', '*')

scatter(x,y, 'r', 'd', 'filled')

Plots: Pie chart

```matlab
%Import .CGmap file using GUI
%Select to import only column with context

context;

x = strmatch('CG', context, 'exact');
y = strmatch('CHG', context, 'exact');
z = strmatch('CHH', context, 'exact');

context2(1) = length(x);     %how many CG context
context2(2) = length(y);     %CHG context
context2(3) = length(z);     %CHH context

pie(context2)
```
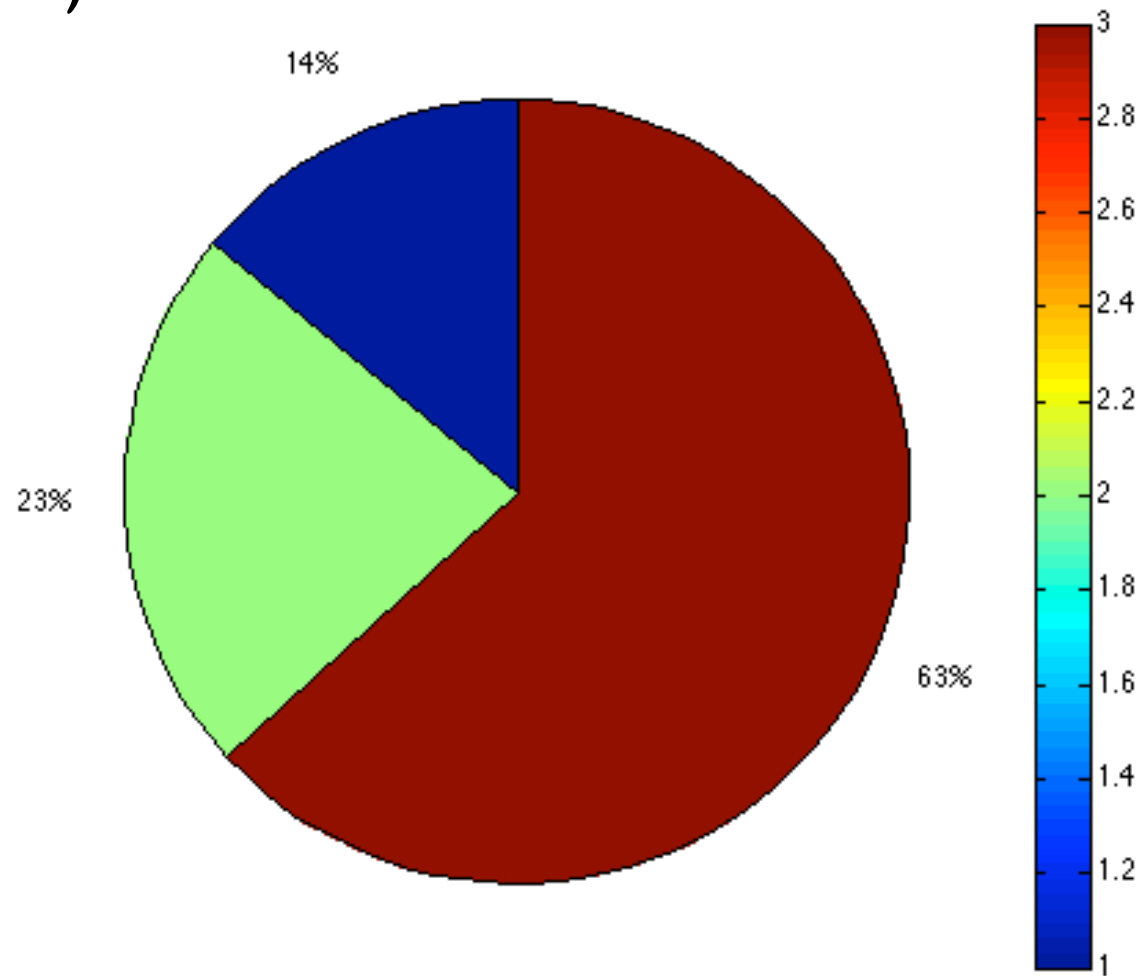
`pie(context2)`

Plots: q-q plot of p-values to uniform distribution 0,1

```matlab
%Create a uniform distribution with
parameters 0 and 1

PD = ProbDistUnivParam('uniform',[0,1]);

p;    %p-values from t-test

figure;
qqplot(p,PD)
```
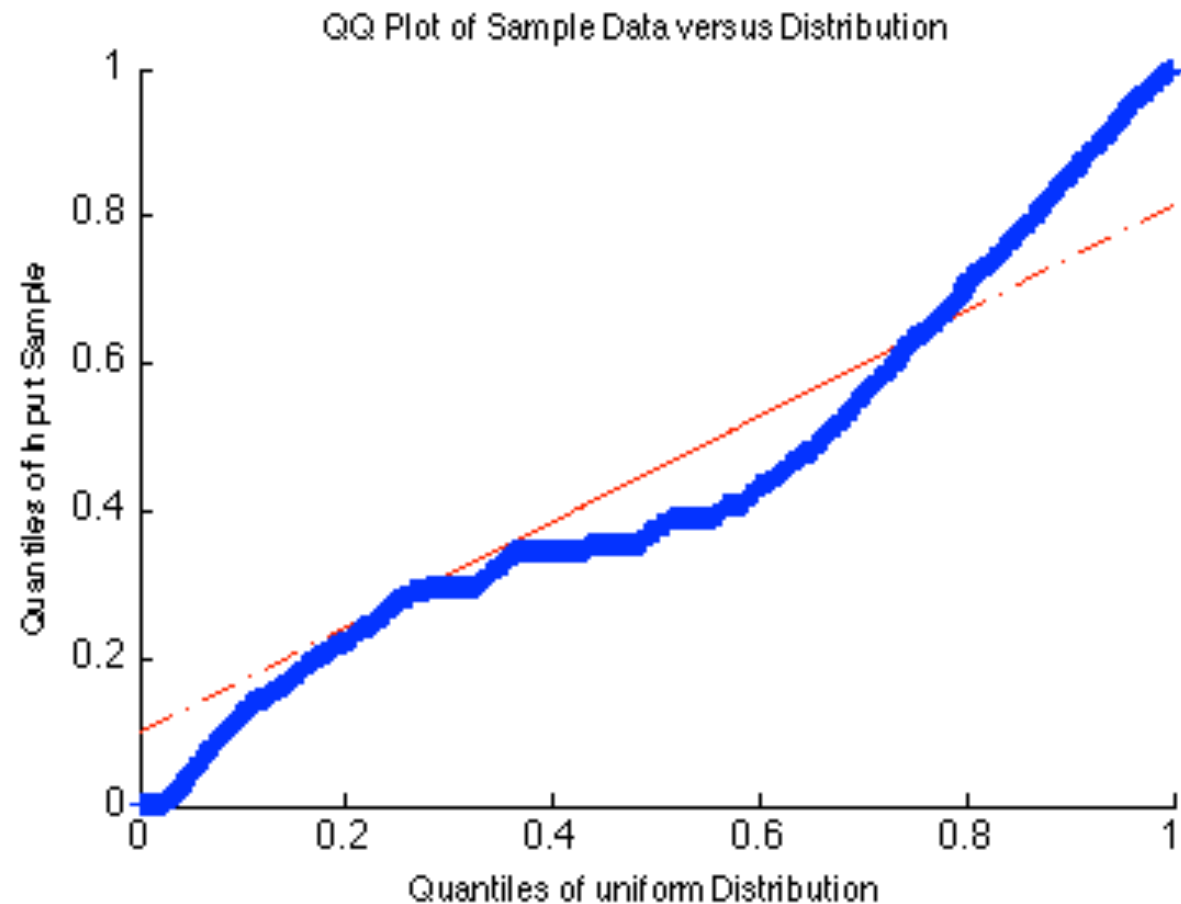
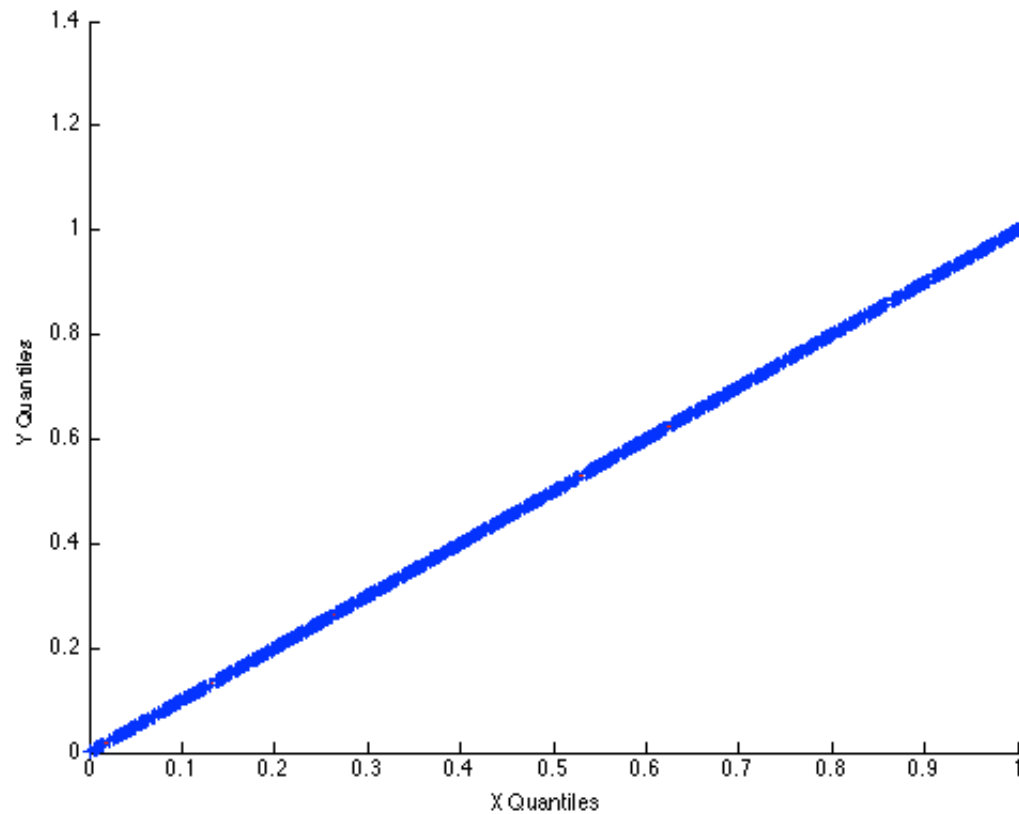## *Plots:* q-q plot of p-values to uniform distribution 0,1

```
figure;
qqplot(p,PD)
```


QQ Plot of Sample Data versus Distribution

Plots: q-q of two p-value distributions

```
figure;
qqplot(p,p_anova)
```

Plots: MATLAB easy plots