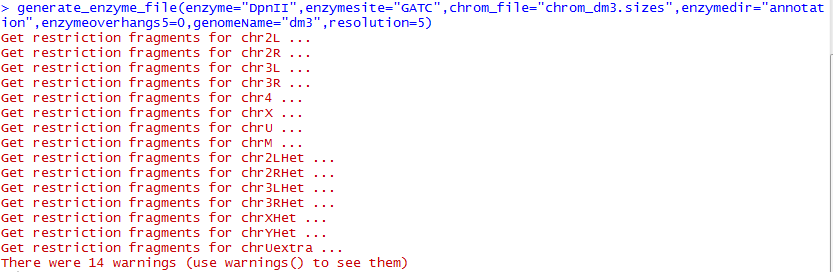
Step 1. Start with our raw fastq datasets.

We can use HBP or HiC-Pro directly to pre-processing our fastq datasets. If we use HBP, we should use following command to generate enzyme site file at first:

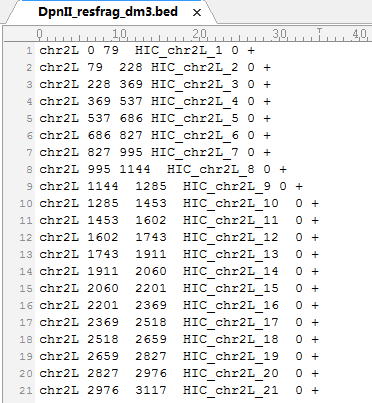
generate\_enzyme\_file(enzyme="DpnII",enzymesite="GATC",chrom\_file="chrom\_dm3.sizes",enzymedir="annotation",enzymeoverhangs5=0,genomeName="dm3",resolution=5)

In this function, “enzyme” is the name of enzyme, “enzymesite” is the restriction enzyme cutting site, “chrom\_file” is the chrom information of genome, the format can be found in the demo, “enzymedir” is the dic of this output file, “enzymeoverhangs5” is the 5' overhangs on the DNA resulted from the cutting, for example, HindIII is 1, and DpnII is 0, “genomeName” is the name of genome, if this genome is not exists in the CRAN or Bioconductor, can use the function generante\_enzyme\_file\_by\_fasta().



After that, we will get the enzyme sites file in the annotation dir. This file is just look like following picture. Then, we can use the following command to processing raw dataset.

run\_hicpro(hicpro\_path = "HiC-Pro",inputfile = "rawdata",configfile = "config-hicpro.txt",outdir = "demoout")

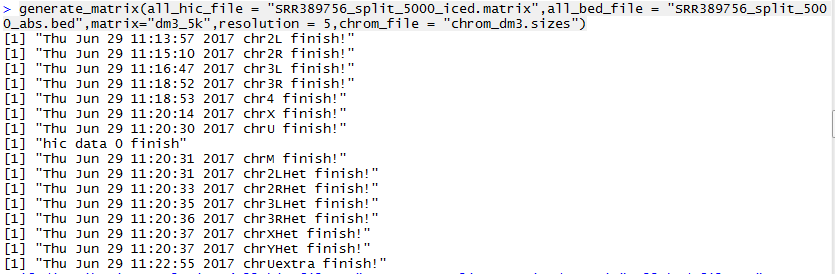


Step 2. Generate interaction frequency matrix.

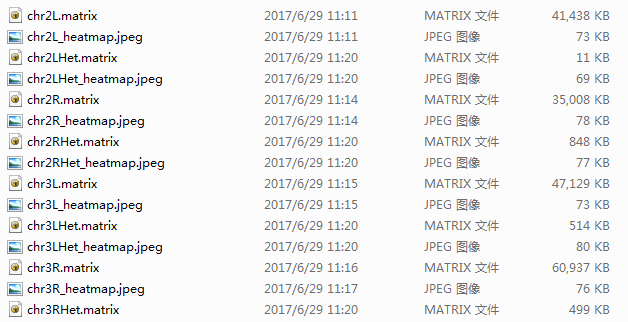
We can get the interaction file from dir “demoout\hic\_results\matrix\SRR389756\_split\iced\5000\SRR389756\_split\_5000\_iced.matrix”, and the index file from dir “demoout\hic\_results\matrix\SRR389756\_split\raw\5000\ SRR389756\_split\_5000\_abs.bed”. And then copy these file to the work dir, and use following command to generate matrix:

generate\_matrix(all\_hic\_file = "SRR389756\_split\_5000\_iced.matrix",all\_bed\_file = "SRR389756\_split\_5000\_abs.bed",matrix="dm3\_5k",resolution = 5,chrom\_file = "chrom\_dm3.sizes")

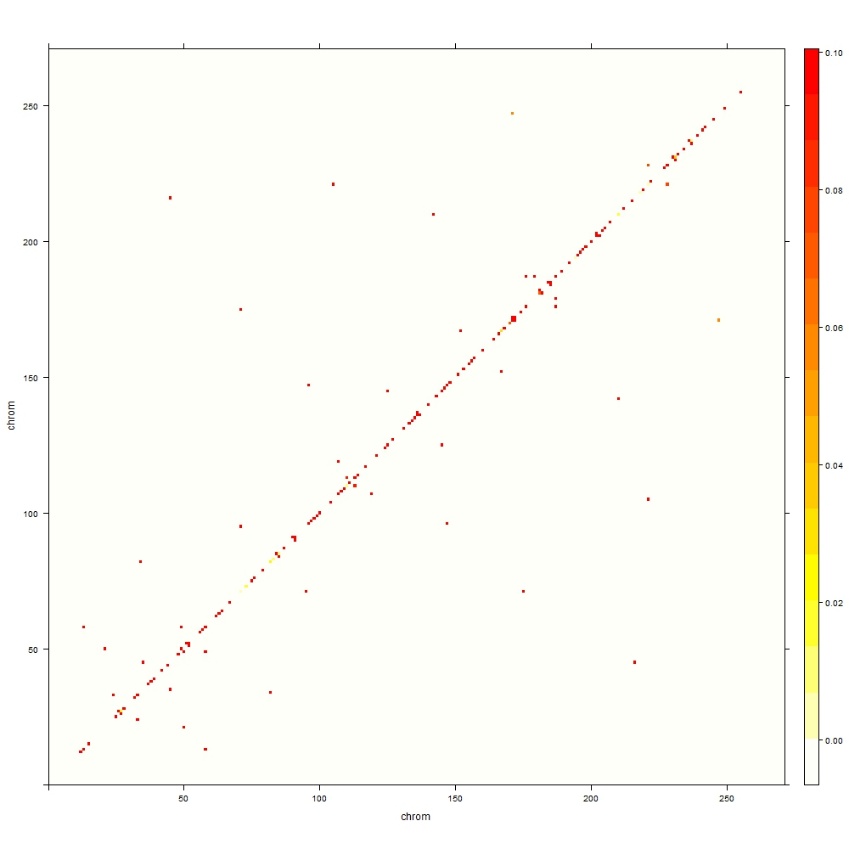
If this function works slowly, we can add parameter “heatmap\_plot = FALSE” to avoid plotting heatmaps, and it will work faster.



The results are just like this picture:



HBP generate interaction frequency matrix file and heatmap at here.



Step 3. Interaction frequency distribution analysis

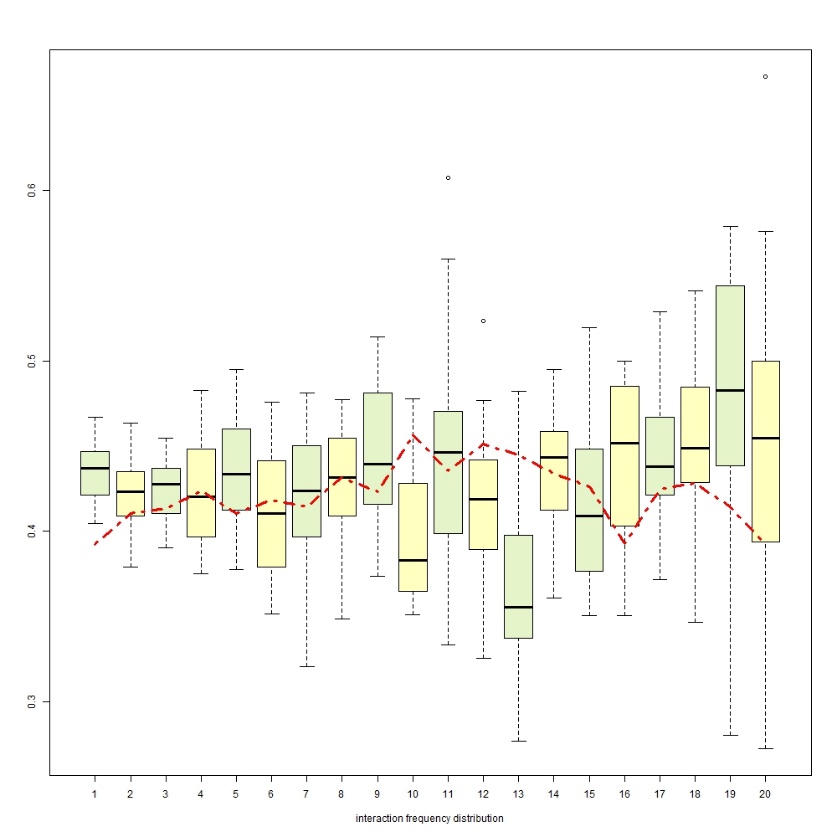
Then we can use the command to make Interaction frequency distribution analysis:

if\_distribution\_analysis(all\_hic\_file = "SRR389756\_split\_5000\_iced.matrix",all\_bed\_file = "SRR389756\_split\_5000\_abs.bed",bedFile = "dm3\_mars.bed",inter\_chromfile = NULL,groupNum = 20,random\_analysis = TRUE,threshold\_percent = 0.005,if\_bin\_number = 20,matrix\_dir = "dm3\_5k",slide\_window = TRUE)

This function have many parameters, and the description can be found in the package, just input the command “?if\_distribution\_analysis”.

The results are looking like following picture.



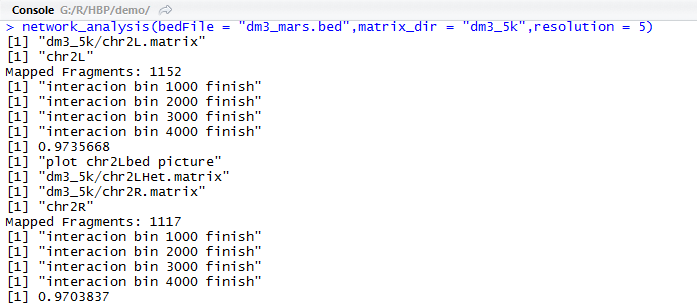


In this picture, the X axis is the interaction strength, and the Y axis is the percent, the yellow and green box plot is random control group, and the red line is the treatment group.

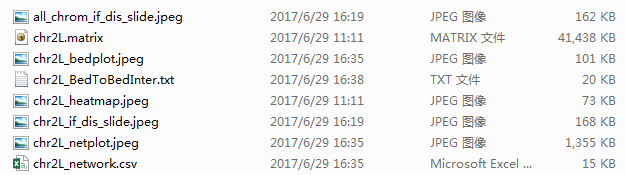
Step 4. Interaction network topological analysis

At here we can use the following command to make interaction network topological analysis:

network\_analysis(bedFile = "dm3\_mars.bed",matrix\_dir = "dm3\_5k",resolution = 5)



After this, we can get some plot and list at the dir “dm3\_5k”:

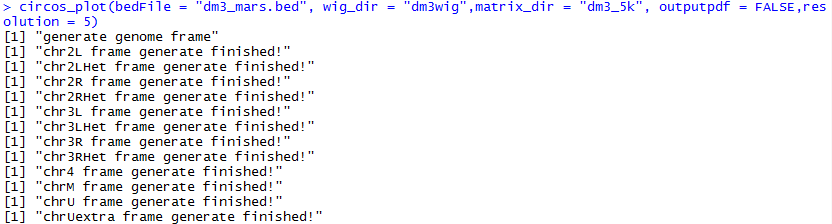


The chr2L\_bedplot.jpeg is the Hi-C heatmap with specific sites. The chr2L\_BedToBedInter.txt is a list contains data about the interaction, which is generated by HBP and used when processing datasets. The chr2L\_netplot.jpeg is topological network plot of this chromosome. If the range is too big, this picture maybe not clearly to investigate, and can be optimize by adjusting parameters. And the chr2L\_network.csv is the list of nodes in this network. This list contains degree, closeness, betweenness, local cluster coefficient, eigenvector centrality and cluster membership information of these nodes.

Step 5. Visualization of interactions and tracks.

We can use following command to plot circos picture of this network:

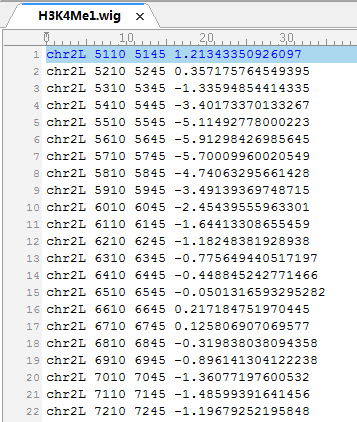
circos\_plot(bedFile = "dm3\_mars.bed", wig\_dir = "dm3wig",matrix\_dir = "dm3\_5k", outputpdf = FALSE,resolution = 5)



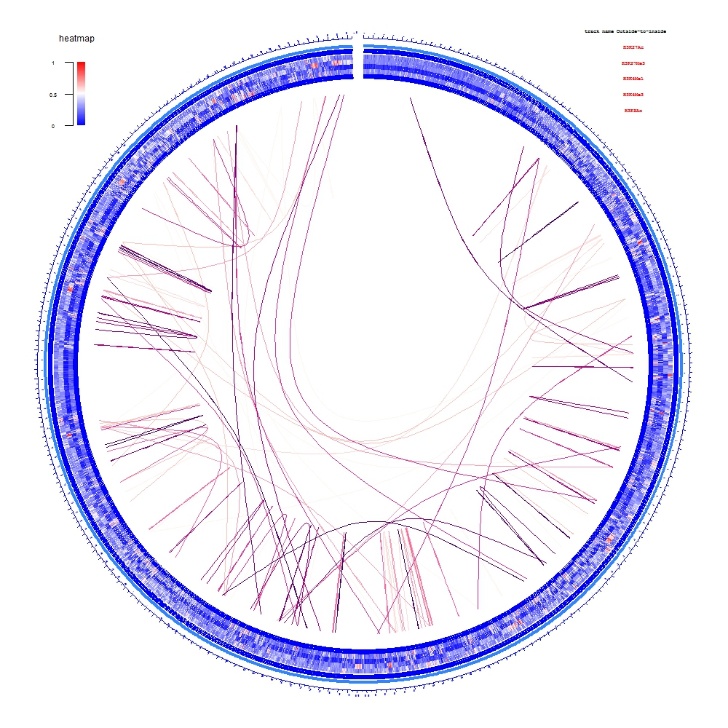
The wig file is stored in the wig\_dir:



And these wig files are downloaded from UCSC:



According to this step, we can get a picture named “\*\_ circos.jpeg”:

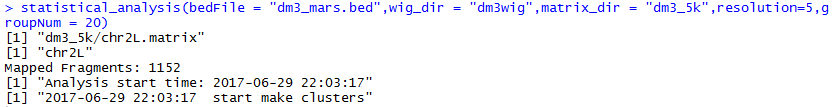


If the range is too big, this picture maybe cannot be see clearly in jpeg format, so we can output pdf format to make it clearly by change the parameter “outputpdf”. And the color of lines and other elements in this picture can be optimized by adjusting other parameters.

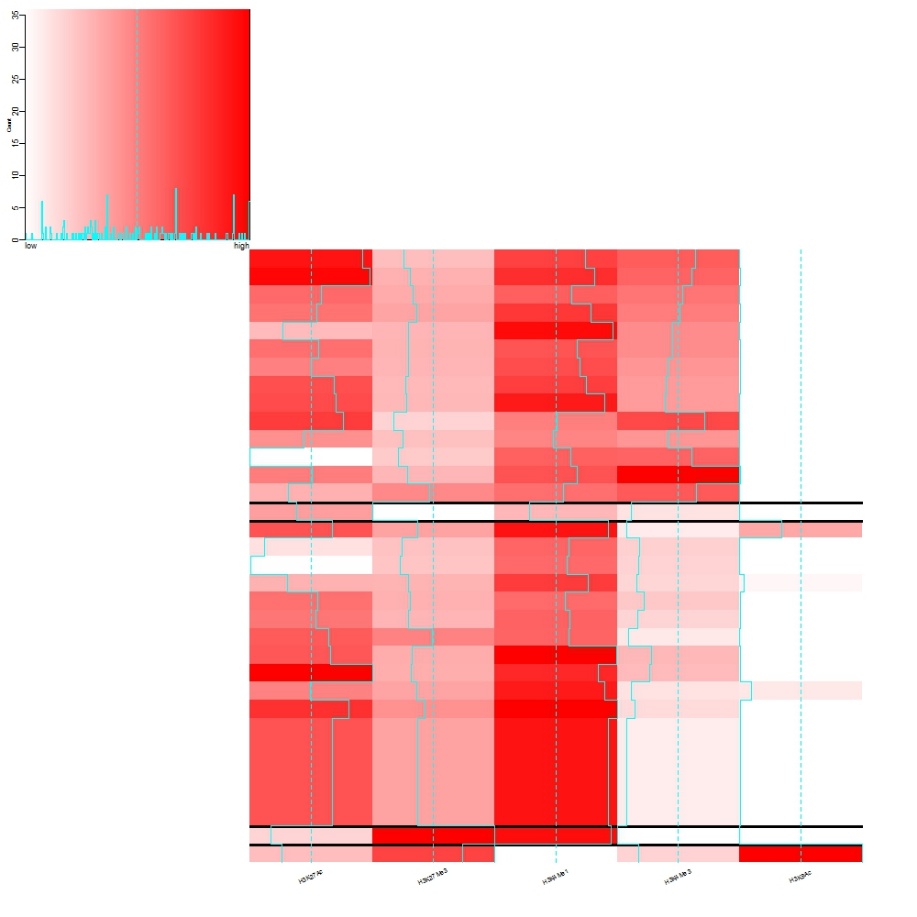
Step 6. Statistical significance tests.

HBP can use following command to calculate statistical significance:

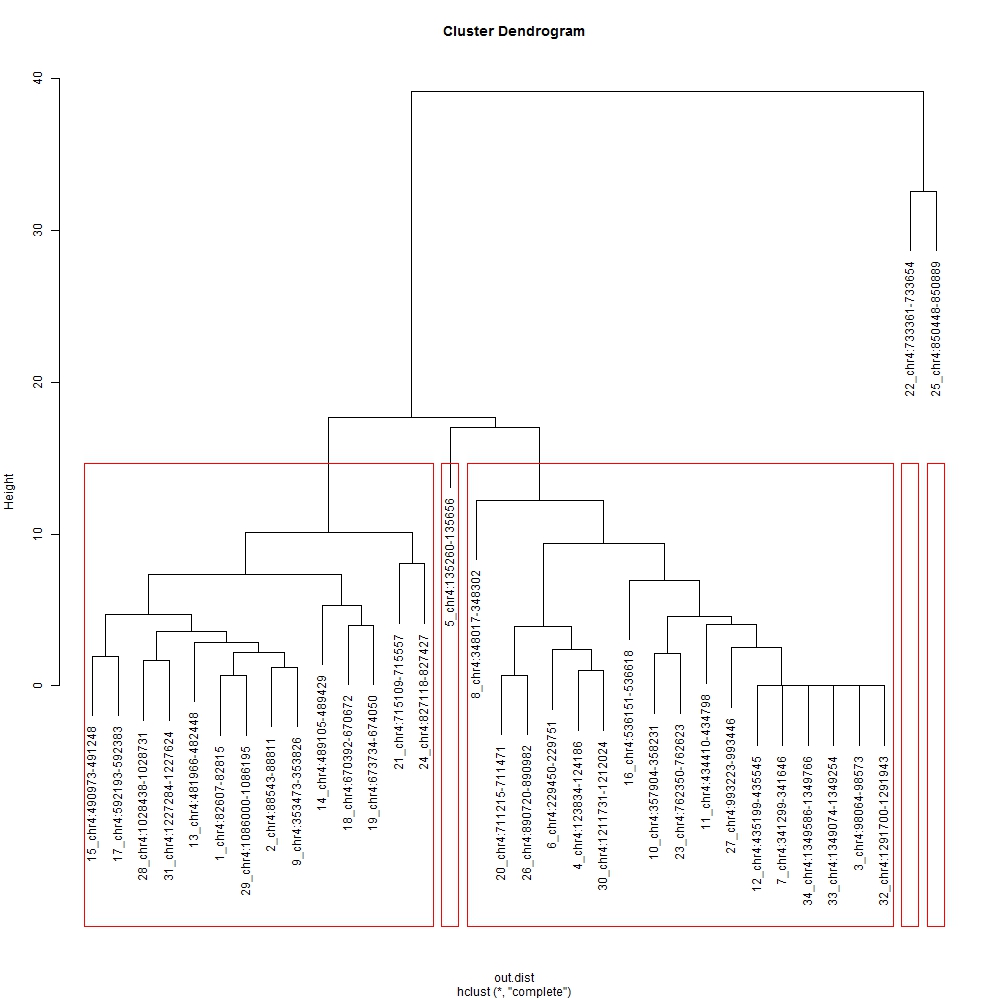
statistical\_analysis(bedFile = "dm3\_mars.bed",wig\_dir = "dm3wig",matrix\_dir = "dm3\_5k",resolution=5)



And we can get several results file from this step. The chr\*\_cluster\_heatmap.jpeg is heatmap of tracks clusters:



The chr\*\_cluster\_tree.jpeg is the cluster tree of these node:



The chr\*\_statistic.txt contains the statistical difference of these interactions:

