

Analyzing Genome-wide Chromatin Accessibility

Pei-Yu Lin

Dr. Pao-Yang Chen's lab

2022.3.16

Outline

- ATAC-seq
- Data processing pipeline
 - Preprocessing
 - Peak calling
- Post-alignment analysis (demo)
 - Accessible regions profiling

Methods for measuring chromatin accessibility

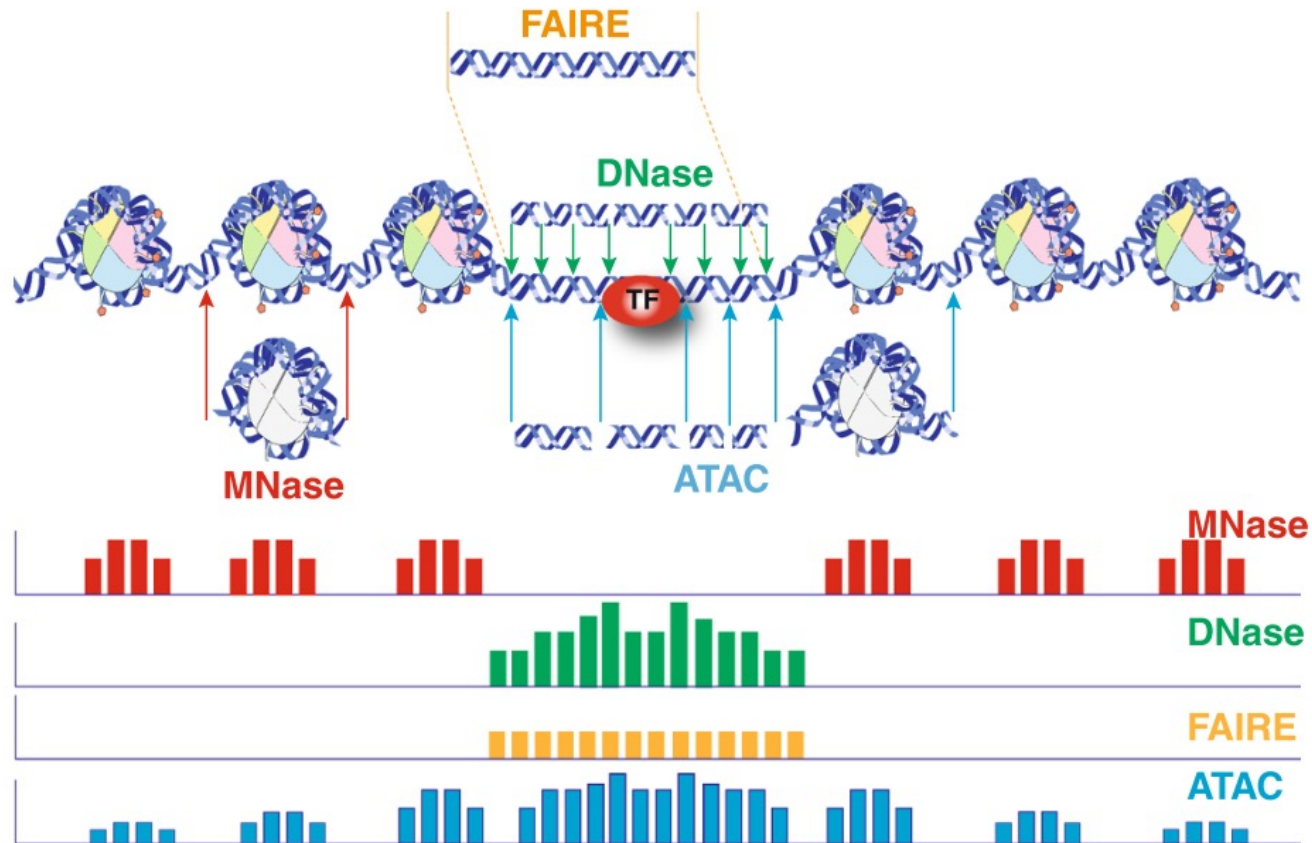
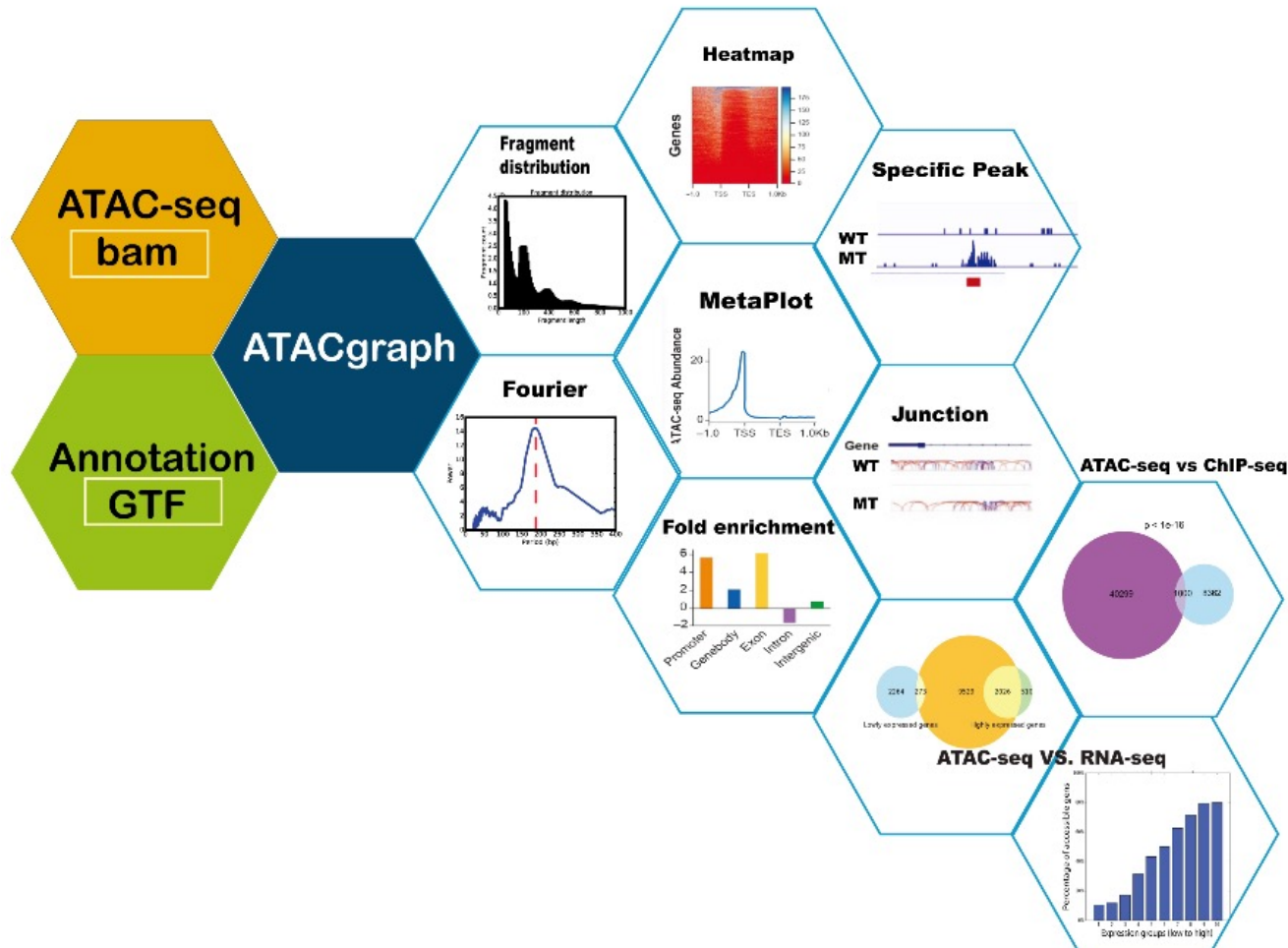


Figure 1 Schematic diagram of current chromatin accessibility assays performed with typical experimental conditions. Representative DNA fragments generated by each assay are shown, with end locations within chromatin defined by colored arrows. Bar diagrams represent data signal obtained from each assay across the entire region. The footprint created by a transcription factor (TF) is shown for ATAC-seq and DNase-seq experiments.

(Tsompana and Buck., 2014)

Chromatin accessibility analysis workflow

All work is done in Linux environment



Data processing – peak calling

1. Remove mitochondria DNA
2. Peak calling

Download and environment setup for ATAC-seq

1. Download ATAC-graph

```
git clone https://github.com/RitataLU/ATACgraph.git  
cd ATACgraph
```

2. Create python2.7 environment for ATAC-graph

```
sh ./ATACgraph/base.txt
```

Example data Human (人類)

- Example:
cd ATACgraph/demo

01_for_data_processing

human genome annotation:

demo_gene.gtf

human gene and promoter bed files:

demo_gene_body_bed6.bed

02_for_data_visualisation

Raw reads bam file:

demo.bam

Raw reads bam index file:

demo.bam.bai

BigWig file:

demo_rmM_peakcall_coverage.bw

Peak location BED file:

demo_rmM_peakcall_peaks.narrowPeak

A genes list of overlapping with peaks locations :

demo_rmM_peakcall_peak_gene_list.txt

03_for_downstream_analysis

Peak location BED file:

demo_rmM_peakcall_peaks.narrowPeak

Remove mitochondria DNA

20-80% sequences in ATAC-seq are from mitochondria genomes

- Remove mitochondria chromosome

```
./script/ATACgraph 00_rmChr demo/demo.bam demo/demo_rmM.bam chrM  
                          Input.bam                  Output.bam          chromosomes
```

```
Remove chrM 69628 reads  
Remove total 69628 out of 71843 (0.969)
```

- Transform GTF file to BED files

```
./script/ATACgraph 02_gtftoBed demo/demo_gene.gtf demo/demo -p 2000  
                          Reference genome.gtf      Output.bed      Promoter region
```


Peak calling

- Peak calling

```
./script/ATACgraph 03_callPeak demo/demo_rmM.bam demo/demo_rmM_peakcall  
demo_gene_body_bed6.bed input.bam Output name  
Gene body.bed
```



- Peak location file

demo_rmM_peakcall_peaks.narrowPeak

- Peak intensity bigwig file

demo_rmM_peakcall_coverage.bw

- A genes list of overlapping with peaks locations

demo_rmM_peakcall_peak_gene_list.txt

Post-alignment analysis

Fragment length distribution and Fast Fourier Transform (FFT)

- Find fragment

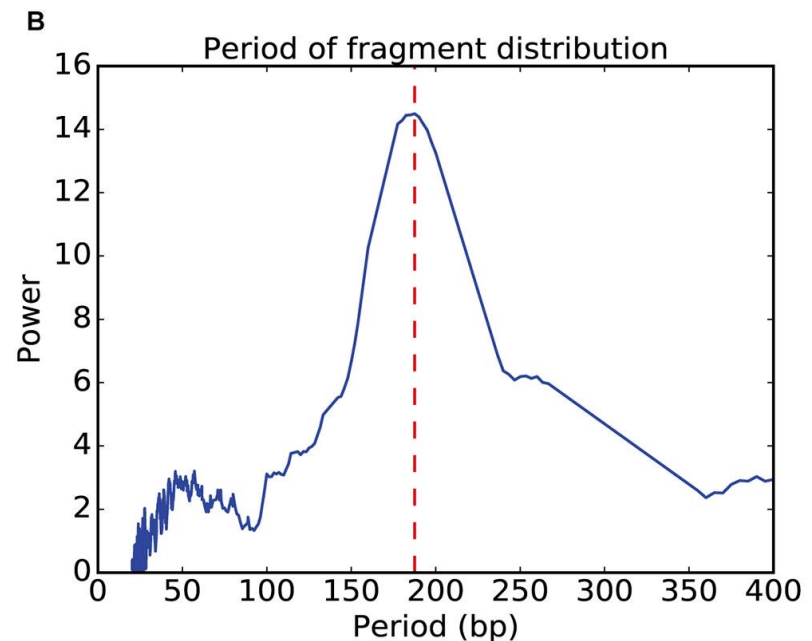
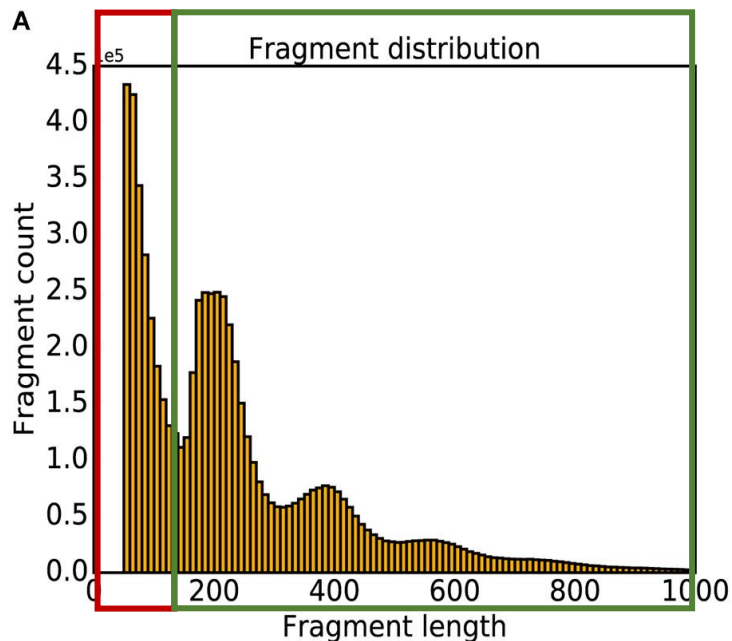
```
./script/ATACgraph 01_calFragDist demo/demo_rmM.bam demo/demo_rmM_fragment  
demo/demo_rmM_FFT
```

input.bam

Output (fragment)

Output (FFT)

NFR (nucleosome free region)



Fragment length distribution and Fast Fourier Transform (FFT)

- Find fragment

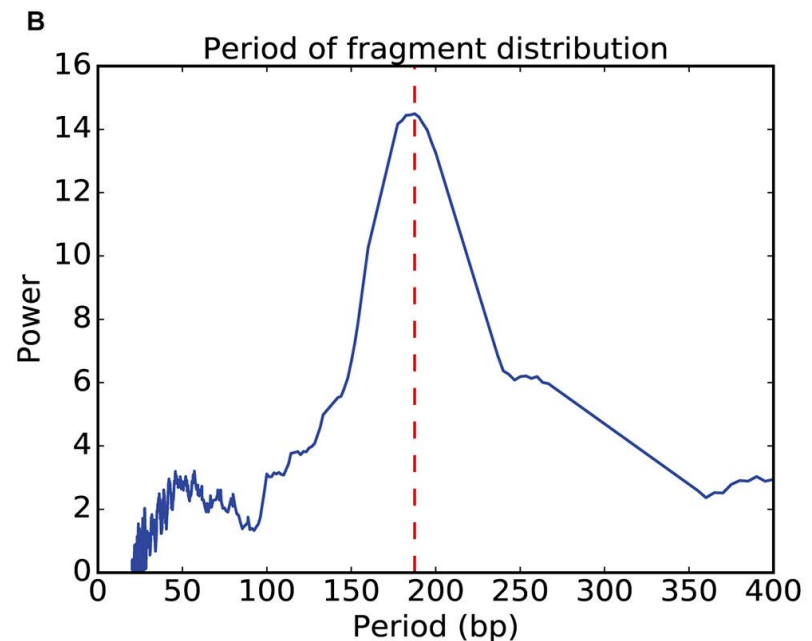
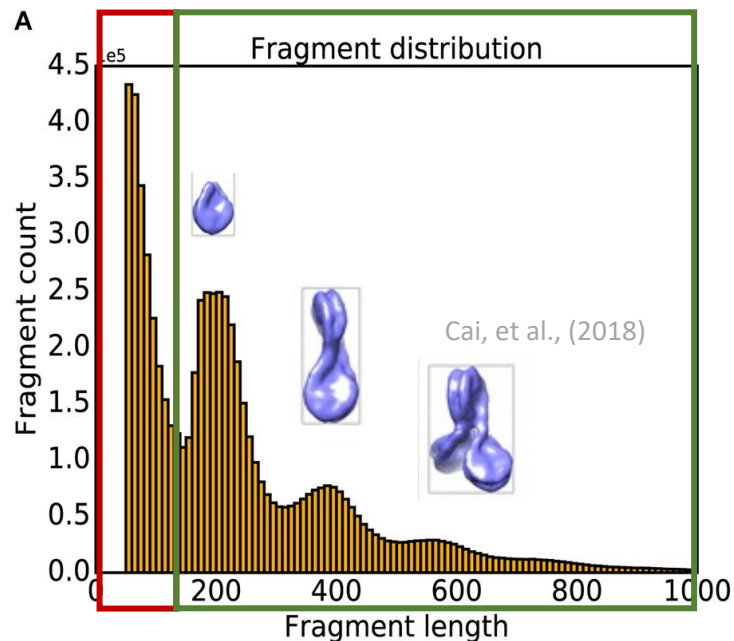
```
./script/ATACgraph 01_calFragDist demo/demo_rmM.bam demo/demo_rmM_fragment  
demo/demo_rmM_FFT
```

input.bam

Output (fragment)

Output (FFT)

NFR (nucleosome free region)



mono-nucleosomes, di-nucleosomes, and tri-nucleosomes

Visualisation of peaks

- Peaks analyses

```
./script/ATACgraph 03_genePlot demo/demo_rmM_peakcall_peaks.narrowPeak  
demo/demo_rmM_peakcall_coverage.bw demo/demo input.narrowpeak
```

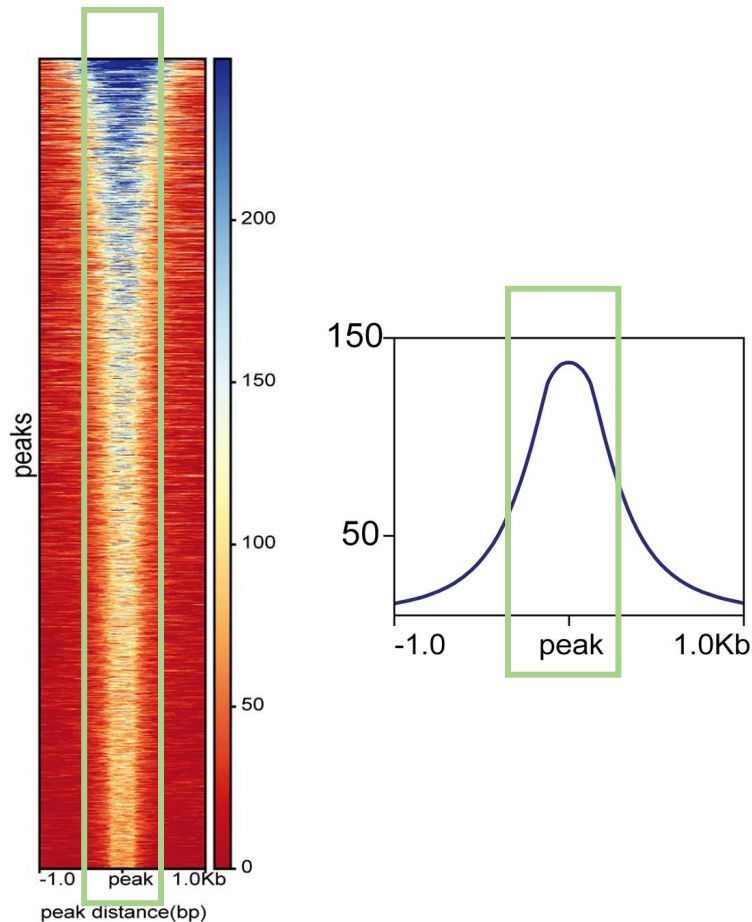
Input.bw **gtf file name**

3 Figures

- The enrichment status of accessible region in genome
Fold_Enrichment.pdf
- The accessibility – or read abundance – around genes
gene_body_heatmap.pdf
- The accessibility – or read abundance – around peaks
Peak_heatmap.pdf

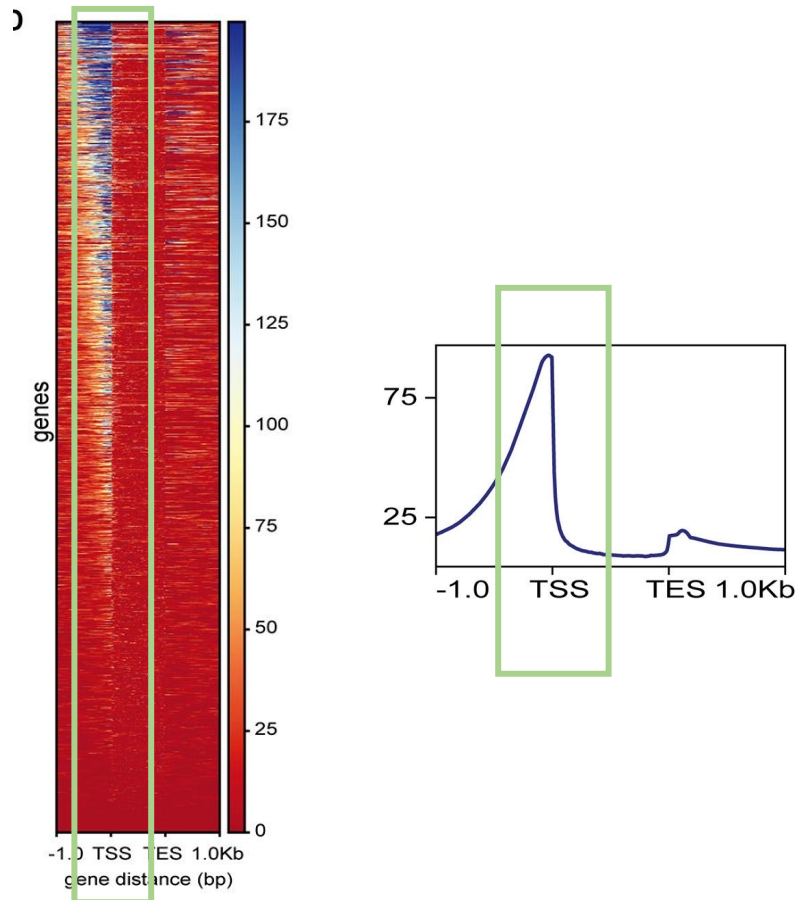
ATAC-seq abundance near the peak regions

- ATAC-seq enriched at the center of the predicted peak locations



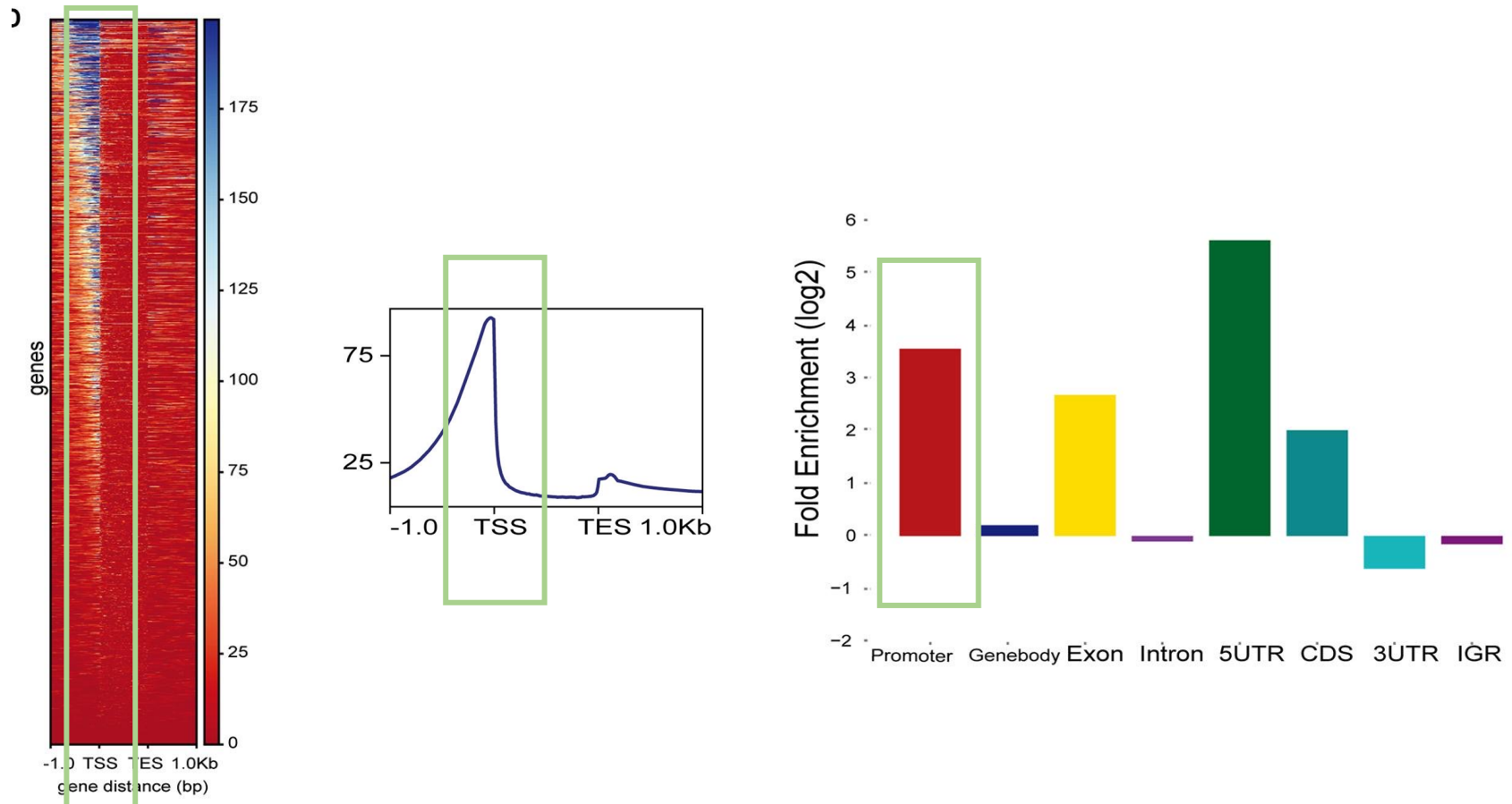
ATAC-seq abundance of the gene body and flanking regions

- The accessible regions are located before the transcription start sites (TSSs) in two-thirds of the genes



ATAC-seq abundance of the gene body and flanking regions

- the ATAC-seq abundance is clearly enriched at promoters close to TSSs, depleted in the gene body, and slightly enriched after the transcription end sites (TESs)



Thank you for listening!