

# General Methods

## 1 Peak Calling

### 1.1 Model-based Analysis of Chip-Seq (MACS)

Taken from [1]. The improved quality of peaks in the Chip-seq data motivates the analysis of the peak calling procedure such as MACS. Specifically, the binding of cis elements with trans factors, which indicates the genes that are directly regulated by those trans factors.

The main points: Chip-seq data exhibit regional biases along the genome due to sequencing and mapping biases, chromatin structure and genome copy number variation. Chip-seq tags are often shifted toward the 3' end. the shift size is unknown. this gives a bi-modal peak pattern for the DNA protein interaction site.

Given a bandwidth (fragment size), MACS slides  $2 \times \text{bandwidth}$  window across the genome to find peaks higher than the background distribution. The background distribution can come from a control experiment. For data with control, MACS linearly scales the control to have the same tag count as the experiment.

Chip seq tag distribution along the DNA is modeled using a Poisson distribution and is considered as a one parameter  $\lambda_{BG}$  distribution from which we can extract suspected peak location. Suspected peaks are called using the background distribution for reads with a p-value of  $10^{-5}$ . due to biases, MACS uses a local  $\lambda$  instead of the uniform  $\lambda_{BG}$

$$\lambda_{local} = \max(\lambda_{BG}, \lambda_{1k}, \lambda_{5k}, \lambda_{10k})$$

where,  $\lambda_{1k}, \lambda_{5k}, \lambda_{10k}$  are  $\lambda$  estimates of a window size  $1k, 5k, 10k$  centered in the peak location in the background (control) sample, or the chip-seq sample when no control sample is used MACS uses  $\lambda_{local}$  to calculate the p-value for each peak and remove false positive due to local bias (that is, peaks significantly under  $\lambda_{BG}$  but not under  $\lambda_{local}$ )

## References

- [1] Yong Zhang, Tao Liu, Clifford A Meyer, Jérôme Eeckhoutte, David S Johnson, Bradley E Bernstein, Chad Nusbaum, Richard M Myers, Myles Brown, Wei Li, et al. Model-based analysis of chip-seq (macs). *Genome Biol*, 9(9):R137, 2008.