Creating 2D Occupancy Plots Using plot2DO

Paula Beati[1], *Razvan V. Chereji*[2]

[1] Instituto de Investigaciones en Ingeniería Genética y Biología Molecular, Consejo Nacional de Investigaciones Científicas y Técnicas, Buenos Aires, Argentina

[2] Division of Developmental Biology, Eunice Kennedy Shriver National Institute for Child Health and Human Development, National Institutes of Health, Bethesda, MD, USA

Abbreviated abstract:

The 2D occupancy (2DO) plot is useful for an initial quality check of MNase-seq data, in order to assess the degree of chromatin digestion, and also for getting insights about the nucleosome organization, as well as about the MNase digestion kinetics.

plot2DO is a flexible tool for evaluating the quality of MNase-seq and MNase-ChIP-seq data, and for visualizing the distribution of nucleosomes near the functional regions of the genome.

Related publication:

– P. Beati et al, Stem Cell Transcriptional Networks. Methods in Molecular Biology, vol 2117 (2020) https://doi.org/10.1007/978-1-0716-0301-7_5

Challenge and approach

Mapping nucleosomes and their epigenetic marks with high accuracy is important for understanding DNA-binding, gene regulation, and all other DNA-related processes in the cell.

Currently, the most commonly used method for determining nucleosome positions is MNase-seq: chromatin is digested with micrococcal nuclease (MNase) and the resulting DNA fragments are subjected to high-throughput sequencing.

Unfortunately, MNase has a strong sequence preference for A/T-rich sequences, and the nucleosomal fragments that result from an MNase-seq experiment contain biases that depend on the degree of MNase digestion. Thus, after the mild digestion of chromatin, a large fraction of the genome is not yet broken into mononucleosomal DNA fragments (~150 bp long) and is still contained in larger oligonucleosome fragments, which are usually not sequenced.

On the other hand, after extensive digestion of chromatin, many nucleosomes occupying A/T-rich sequences are overdigested and lost from the sample of mononucleosomal (~150 bp long) fragments.

Challenge and approach (cont.)

Therefore, MNase-seq experiments require careful control of the level of digestion, and the variable degree of digestion must always be considered when multiple samples are compared and used to draw conclusions from the differences observed among the samples.

plot2DO is useful for plotting the coverage of sequencing reads of various sizes at different genomic loci. The 2D occupancy (2DO) plot is useful not only for the initial quality check of MNase-seq data, in order to assess the degree of digestion, but also for getting insights about the nucleosome organization near functional regions of the genome, as well as about the MNase digestion kinetics.

plot2DO is based on ggplot2, and Bioconductor libraries AnnotationHub, biomaRt, GenomicFeatures, GenomicRanges, GenomeInfoDb, Rsamtools, rtracklayer and it is available from https://github.com/rchereji/plot2DO

Techniques and Methods

Algorithm

Inputs: reads aligned to the reference genome, the genome version, the read lengths range, the reference points to be aligned, the type of distribution to plot (occupancy or dyad), the length of upstream and downstream regions to be plotted

Do: Read transcript annotations for the genome of interest
Create genomic intervals centered of specified reference points (e.g. TSS, TTS)
Select MNase-seq reads of different lengths in genomic intervals of interest
Compute distributions of MNase-seq read dyads and occupancy/coverage, and plot heat maps

Heatmap values computation (coverage for reads aligned to reference points)

Occupancy is calculated for each read length in a given range, weighted by a normalization factor by chromosome. The normalization factor by chromosome is defined as the reciprocal of the mean occupancy for that chromosome. Then, these occupancy values are aligned to the specified reference points and the mean of coverage by length and the position (relative to each reference point) is reported.

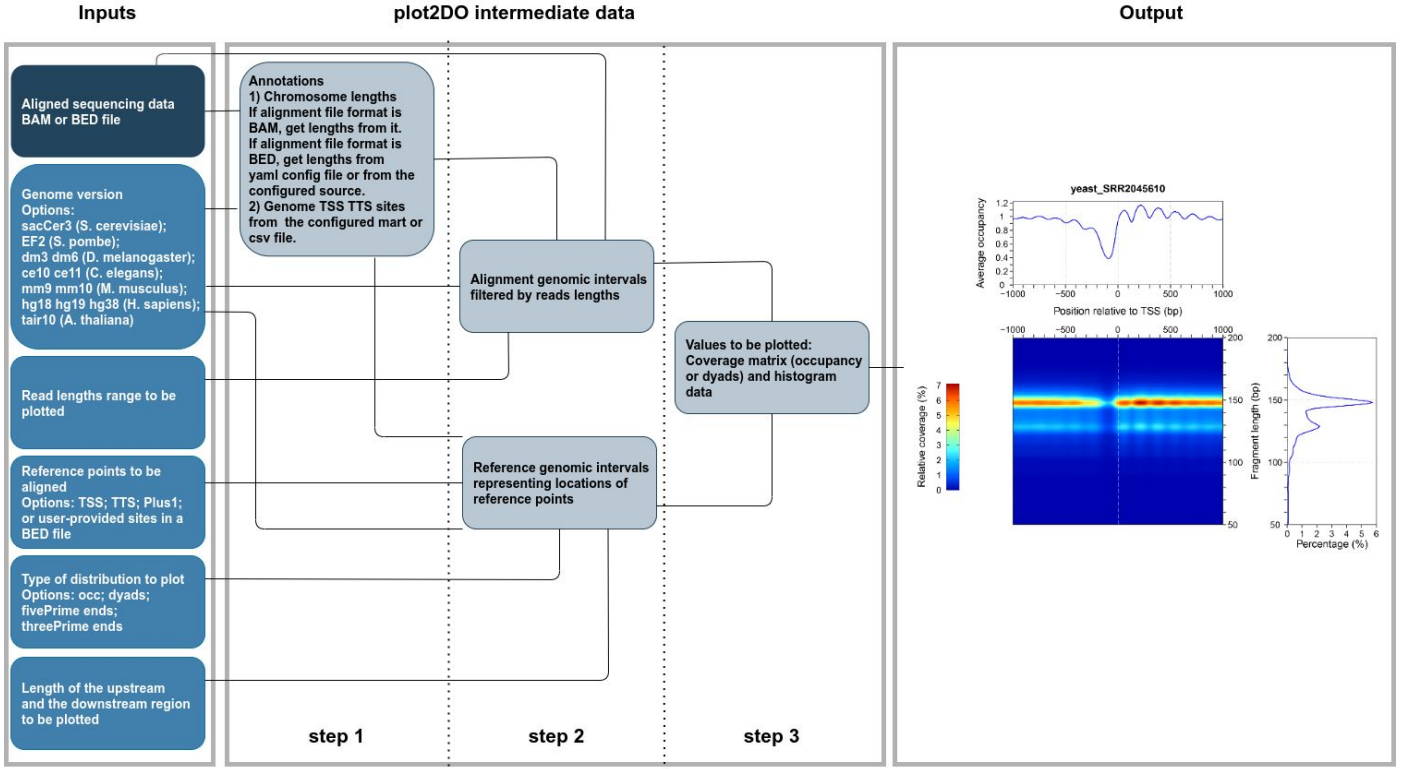

Figure 1 | plot2DO data processing stages

Results

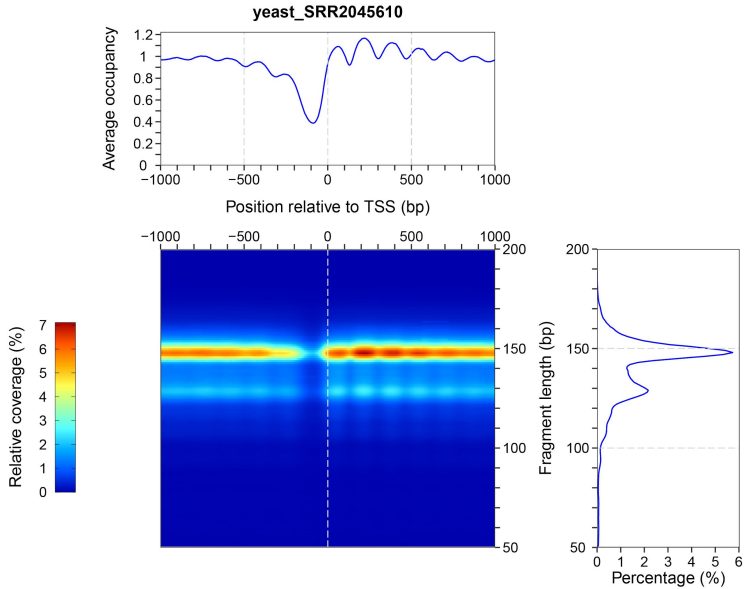

Figure 2 | plot2DO output example

From the panels produced by plot2DO, we can obtain the following information:

Top panel shows the traditional average occupancy plot, obtained by stacking all the sequencing reads and averaging the obtained coverage for all genes, aligned at their reference point (e.g. TSS – transcription start site). We can observe the nucleosome depleted region (NDR) that forms at the gene promoters, and the degree of nucleosome phasing how consistent this regular nucleosome organization is near the reference points, among all genes.

Results (cont.)

Right panel shows the degree of chromatin digestion. We should observe most DNA fragments having a length of about 147 bp, the size of a nucleosome core particle, but maybe some of the fragments correspond to overdigested nucleosomes, which would have lost some bp of DNA from each end, and would show a reduced footprint.

2DO plot contains all the information of the previous 1D plots, in a disentangled way.
We can now distinguish the contribution to occupancy of DNA fragments of different sizes occupying the same positions in different genes or cells, which is not possible from the typical 1D occupancy plot.
It is also possible to observe the distribution near promoters of all fragments corresponding to a given size, which is not possible from the usual 1D fragment length distribution.

Conclusions

The plots generated by plot2DO are particularly useful for investigating the level of digestion and the effect of MNase on the subset of nucleosomes that are obtained in MNase-seq experiments.

Apart from MNase-seq data analysis, plot2DO can be useful to investigate DNA fragments obtained in other types of genome-wide experiments involving chromatin fragmentation, such as ChIP-seq, ATAC-seq, and chemical cleavage mapping.

plot2DO can be used to analyze the amount of DNA that is obtained from different regions of the genome, and the corresponding distribution of fragment sizes for each region, independent of the protocol that is used to generate the short DNA fragments.

We recommend that in MNase-seq, MNase-ChIP-seq, and perhaps in other genome-wide techniques that break the chromatin to produce short DNA fragments, plot2DO should be the first plot, in order to assess the level of chromatin fragmentation and the quality of the genomic data.