ChIP-seq data processing

Data was processed using the *ChIP-nf* (https://github.com/guigolab/chip-nf) Nextflow[1] pipeline. ChIP-seq reads were aligned to the human genome assembly (GRCh38) using the GEM[2] mapping software, allowing up to two mismatches. Only alignments for reads mapping to ten or fewer loci were reported. Duplicated reads were removed using Picard (http://broadinstitute.github.io/picard/). We obtained peak-calling BED files, and fold-change and p-value bigWig files, for individual replicates by running MACS2[3]. No shifting model was built. Instead, fragment length was set to 250 bp and was used to extend each read towards the 3' end (using the *--extsize* option). Peak calling was also performed jointly on biological replicates with Zerone[4], and passed the filter for all pairs of replicates (*advice: accept discretization*). To check library complexity, we computed the fraction of non-redundant mapped reads[5] (recommended threshold: NRF ≥ 0.8) for each ChIP-seq experiment, and found a minimum NRF value of 0.92 across all ChIP-seq experiments. Additionally, to evaluate the global ChIP enrichment, we computed the fraction of reads in peaks (Landt et al., 2012) (recommended threshold: FRiP ≥ 0.01), and found a minimum FRiP value of 0.05 across all ChIP-seq experiments.

References

1 Di Tommaso, P. *et al.* Nextflow enables reproducible computational workflows. *Nature biotechnology* **35**, 316-319, doi:10.1038/nbt.3820 (2017).

2 Marco-Sola, S., Sammeth, M., Guigo, R. & Ribeca, P. The GEM mapper: fast, accurate and versatile alignment by filtration. *Nature methods* **9**, 1185-1188, doi:10.1038/nmeth.2221 (2012).

3 Zhang, Y. *et al.* Model-based analysis of ChIP-Seq (MACS). *Genome biology* **9**, R137, doi:10.1186/gb-2008-9-9-r137 (2008).

4 Cusco, P. & Filion, G. J. Zerone: a ChIP-seq discretizer for multiple replicates with built-in quality control. *Bioinformatics* **32**, 2896-2902, doi:10.1093/bioinformatics/btw336 (2016).

5 Landt, S. G. *et al.* ChIP-seq guidelines and practices of the ENCODE and modENCODE consortia. *Genome research* **22**, 1813-1831, doi:10.1101/gr.136184.111 (2012).