ChIP-Seq Analysis Report
Run 1674
COU-BIF-P2
Genome hg19

Bioinformatics Core Facility, IRCM

Alexis Blanchet-Cohen, Bioinformatics analyst
alexis.blanchet-cohen@ircm.qc.ca

May 11, 2014

Contents

1 Sample names and read statistics

Table 1 contains the sample names and the read statistics for the raw reads coming off the sequencer. Detailed statistics on the quality of the raw reads were computed with FastQC [1], and can be found in the *fastqc* folder. After examining the quality of the reads, no trimming was deemed necessary.

Table 1: Read statistics

Name	Run	Run Type	Library Type	Number of Reads	Number of Cycles	Duplicate
DF_3A-Input.G672E	1674	PAIRED_END	ChIP-Seq	41,834,711	50	6.456 %
DF_3A-Input.M852V	1674	PAIRED_END	ChIP-Seq	44,001,716	50	6.13 %
DF_3A-Input	1674	PAIRED_END	ChIP-Seq	36,100,153	50	6.602 %
DF_3A-IP.M852V	1674	PAIRED_END	ChIP-Seq	39,326,721	50	11.015 %
DF_3A-IP	1674	PAIRED_END	ChIP-Seq	36,493,637	50	24.649 %
DF_3A-IP.G672E	1674	PAIRED_END	ChIP-Seq	38,060,476	50	22.407 %

2 Alignment with Bowtie2

The reads were aligned to the hg19 genome with Bowtie2 [3]. The alignment files (BAM files) outputted by Bowtie2 can be viewed with the open source software IGV (`http://www.broadinstitute.org/software/igv/download`). The BAM files can be found in the folder *Analysis/bowtie2*

3 UCSC track files

Since the BAM files are large files, they were converted to smaller bigWig files, which can be uploaded faster in IGV or the UCSC Genome Browser. The bigWig files are less detailed than the BAM files. Any information about the quality of the alignments or mutations in the aligned reads, for example, is not present in the bigWig files. The bigWig files can be found in the folder *Analysis/bigWig*

4 Peak calling with MACS2

4.1 callpeak

Peaks were called using the MACS2 [5] function callpeak. The results can be found in the folder *Analysis/macs/callpeak*

4.2 bdgdiff

The MACS2 function bdgdiff was used to find the differential peaks between the treatment conditions, after the treatment versus control (IP versus input) peaks were found. The MACS2 bdgdiff results are in the folder *Analysis/bdgdiff*.

5 HOMER

Peaks called with MACS were annotated with HOMER [2], using RefSeq annotations. The HOMER results can be found in the folder *Analysis/homer*.

5.1 Gene and genome ontology

Gene ontology and genome ontology analyses were also carried out with HOMER. The gene ontology and genome ontology results are found in the subfolders *gene_ontology* and *genome_ontology* for each peak comparison.

5.2 Motifs

Motif discovery was carried out with HOMER. The motif results are in the subfolders *gene_ontology* and *genome_ontology* for each peak comparison.

6 Command logs

All the command used during the analysis were logged in the folder *Analysis/scripts*.

Appendices

A Software versions

- Bowtie 2.1.0
- MACS 2.0.10
- HOMER 4.5.0

References

[1] S. Andrews. Fastqc: A quality control tool for high throughput sequence data. *http://www.bioinformatics.babraham.ac.uk/projects/fastqc/*, 2012.

[2] S. Heinz, C. Benner, N. Spann, E. Bertolino, and et al. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and b cell identities. *Mol Cell*, 4(38):576–589, 2010.

[3] B. Langmead and S. Salzberg. Fast gapped-read alignment with bowtie 2. *Nature Methods*, 9:357–359, 2012.

[4] M. Lohse, A. M. Bolger, A. Nagel, A. R. Fernie, J.E. Lunn, M. Stitt, and B. Usadel. Robina: a user-friendly, integrated software solution for rna-seq-based transcriptomics. *Nucleic Acids Res.*, 40, 2012.

[5] Y. Zhang, T. Liu, C.A. Meyer, J. Eeckhoute, D.S. Johnson, C. Bernstein, B.E. Nusbaum, R.M. Myers, M. Brown, W. Li, and X.S. Liu. Model-based analysis of chip-seq (macs). *Genome Biology*, 2008.