Methods (from Teytelman L, et al. 2019)

Lenny Teytelman

Abstract

This includes portions of the methods section from:

Teytelman L, Özaydın B, Zill O, Lefrançois P, Snyder M, Rine J, et al. (2009) Impact of Chromatin Structures on DNA Processing for Genomic Analyses. PLoS ONE 4(8): e6700. https://doi.org/10.1371/journal.pone.0006700

The goal of the exercise is to note gaps and missing details. What would/should this M&M section look like today?

Citation: Lenny Teytelman Methods (from Teytelman L, et al. 2019). **protocols.io** dx.doi.org/10.17504/protocols.io.r63d9gn
Published: 31 Jul 2018

Document

Exerpts from: http://journals.plos.org/plosone/article?id=10.1371/journal.pone.0006700#s4

Materials and Methods

Input-Seq datasets, mapping, and filtering

The *S. cerevisiae* formaldehyde cross-linked sheared input samples were prepared as described [28]. *S. cerevisiae* samples were sequenced using the Illumina Genome Analyzer. *S. cerevisiae/S. bayanus* hybrid diploids were generated by crossing *S. cerevisiae* strain W303 to an *S. bayanus* strain derived from the type strain, CBS 7001 (see [29] for details). The *S. cerevisiae/S. bayanus* input samples were prepared for ChIP analysis by formaldehyde cross-linking and sonication as previously described [30]. ChIP-sequencing libraries were prepared as per the Illumina paired-end library protocol, with modifications as per [28], [31]. Following adapter ligation, 500 bp library inserts were selected on a 2% agarose gel. The genomic library was prepared from the parent *S. bayanus* strain. Libraries were sequenced by 36 bp paired-end reads on the Illumina Genome Analyzer II. The *S. cerevisiae* genomic reads are from the *Saccharomyces* Genome Resequencing Project, and included only the S288C *S. cerevisiae*strain [11], also sequenced using the Illumina Genome Analyzer.

The reads were mapped to the *S. cerevisiae* genome using the MAQ software [32]. Due to poly-A sequencing bias of the Illumina Genome Analyzer, we excluded all reads mapping within 50 base-pairs of a run of 10 or more consecutive adenines or consecutive thymines. The sequences in the rDNA locus (chromosome XII, positions 430,000–520,000) were also not analyzed, as the published SGD genome assembly includes only one of the numerous genomic copies of the rDNA, resulting in artificially inflated coverage of this locus.

All sequence reads from the *S. cerevisiae/S. bayanus* input-Seq and the *S. bayanus* genomic sequencing have been deposited in the NCBI Short Read Archive under accession SRP000997.

Input/genomic normalization

Every base of the genome was assigned the total number of sequence reads overlapping it, separately for the input and genomic sequence reads. Subsequent normalization and analysis, with the exception of transcription stop site coverage analysis, was performed on median read coverage across 100 bp windows, sliding along each chromosome in 50 bp steps. The median input coverage of each 100 bp interval was divided by the median genomic coverage for the same window. All of the sequence coverage analyses, with the exception of 3′ transcription ends, were done on the 100 bp windows.

...

Comparison of input-Seq experiments

For each 100 bp window with sequence reads from the *S. cerevisiae* input-Seq, median read counts were calculated for each of the four *S. cerevisiae/S. bayanus* hybrid input-Seq experiments. Scatter plots and correlation coefficients were plotted for all pairs of the hybrid input-Seq datasets. Subsequently, for each base of the genome, sequence read counts were summed from all four hybrid input-Seq mapped results. The combined *S. cerevisiae/S. bayanus* read counts were then used to calculate median coverage for the 100 bp windows, and then compared to the *S. cerevisiae* input-Seq coverage in a scatter plot. Comparisons of the genomic coverage to the hybrid input-Seq coverage were based on the summed counts.

S. bayanus telomeric coverage

To identify putative *S. bayanus* telomeric sequence, we used NCBI BLAST [33] without repeat masking (-F F), searching with all of the *S. cerevisiae* annotated telomeric DNA against the Washington University *S. bayanus* assembly [34]. We used e-value cutoff of 0.1 and only accepted matches that were within 5,000 base pairs of a contig end. “Telomeric” coverage was calculated within 500 flanking base pairs of the BLAST matches.

...

Statistical Analyses

All statistical tests were performed using R [36].