SRBrowse Manual v0.1

July 1, 2020

1 Introduction

SRBrowse is designed to make browsing methylome data at single-read resolution easy and straightforward, with a variety of visualisation options and analyses available. The tool optimises memory usage by indexing compact read files (containing only relevant methylation information) during alignment, and serving to the user only the sections of the genome that are displayed. The tool serves the full read data on every refresh, and so it is less suited for averaging data of large regions (e.g. > 50kb). The tool is intended to run on a local machine rather than being deployed on a remote server; accessing it remotely may result in latency when viewing large regions.

2 Usage

2.1 Browsing

The genome browser is accessible from the main page or "Browse" link from the top navigation menu. After the assembly is chosen, methylome tracks and annotations associated with the assembly may be added by clicking on their labels in the left navigation sidebar.

2.1.1 Methylation signal

The basic track view displays an averaged methylation signal, separated into contexts (CG, CHN, CHG and CHH), and stranded according to context. Methylation contexts in the reference genome are drawn as squares on the middle line separating the two strands. Contexts may be hidden by clicking on the context name on the left navigation sidebar. Navigating between regions is possible either through the zoom and movement buttons above the genome browser, by inputting the sequence name (e.g. chromosome or scaffold label) and position manually into the browser navigation, or by searching for specific annotation accessions in the loaded annotation track.

2.1.2 Reads display

It is possible to view reads data (and methylated Cs within reads) by selecting the "Reads" option from the dropdown on the right-hand side of the browser navigation (which appears as "Methylation"). This dropdown selects the display mode, and includes all other display modes available. The reads mode displays all reads overlapping the viewable region, and draws Cs within reads associated with a particular context in the reference genome as boxes (with the same colour as the context's colour code). Reads are also drawn either above or below the middle line according to the strand to which they aligned. It is possible to manipulate the visual display of reads by opening the settings menu (the right-most button): track size (the height of each track in the browser), read size (the pixel height of individual reads), and minimum Cs per read.

2.1.3 Region display

The other display modes available from the display menu (other than Methylation and Reads) relate to region features. In this mode, the signal shown relates to discrete regions which have been identified in the

current displayed region, based on the parameters indicated in the display settings menu (under the subtitle "Region selection"). Regions are matched according to the currently displayed contexts. Thus, it is possible to alter the region parameters to match the current context, or to explore other parameter ranges. Similar to the methylation signal, the region feature is also stranded and drawn above or below the middle line accordingly.

2.2 Analysis

Analyses are initiated either by selecting a single region of interest, or by clicking on an annotated element, and is performed on the currently displayed methylomes. The analysis defines regions genome-wide according to the annotation and region parameters, collects reads from these regions, and calculates region features according to these reads. Once initiated, the progress of the analysis can be viewed in the "Activity" menu. When complete, a CSV file of the analysis can be downloaded from the "Analyses" page (accessible from the top navigation). Analyses are identified by their parameters and selected tracks. Clicking on the "Export" (graph icon) button downloads a CSV file containing a list of matched regions from the selected annotated element type, with features as columns.

2.3 Adding datasets

2.3.1 Indexing genomes

Assemblies can be indexed from the datasets page (accessible through the top navigation) by clicking on the "Add Assembly" button. This will open a form with a user-defined label, and URL of the assembly FASTA file. When the form is submitted, the tool will begin to convert the assembly for BS-seq alignment and when completed run bowtie2-build (Bowtie2's genome indexing utility).

2.3.2 Adding tracks (methylomes)

Once an assembly is indexed, BS-seq data may be added to the tool using the "Add Track" button on the datasets page. Data may be added from the local computer or from NCBI SRA (via either EBI or the fastq-dump utility). Datasets added from SRA are aligned on-the-fly once the download begins. The form includes basic options for the alignment, including the maximum number of mismatches, duplicates and multimappers. Track adding occurs as a background process and can be monitored from the top left "Activity" menu, which provides information on reads processed, alignment percentage, and an estimate to task completion.