# MuSERA User Manual

Vahid Jalili

# Contents

# List of Figures



# List of Tables

# 1 Introduction

MuSERA is a novel, efficient and easy-to-use advanced graphical tool of broad utility to analyze replicates of next-generation sequencing (NGS) experiments whose analysis involves the detection of enriched regions (ERs), as for example in ChIP-seq or DNase-seq datasets. It efficiently implements, extends and generalizes the original method presented in [1] to combine ER evidence across replicates in order to increase the statistical significance of the ERs detected in the NGS experiment; it assigns ERs to different sets, and in addition provides analysis features that allow performing further assessments and functional analyses on the identified ERs. Through its intuitive graphical interface, it provides several graphic displays that help the user in gaining a deeper insight and biological evaluation of the analysis results. MuSERA allows the annotation of samples with user-defined genomic features and the visualization of results in an integrated genome browser. Furthermore, it provides a rich set of quantitative evaluations and interactive graphical displays for functional analysis and global correlation assessment of enriched regions, as well as for nearest enriched region distance distribution, which allow in-depth investigation of each of the determined ER sets in the genomic context and greatly help the understanding and biological interpretation of results.

MuSERA bins distances based on a user-modifiable window size, shows results on tables and plots (supporting user-friendly zoom and pan), and allows operations to be applied on user-selected chromosomes. How to use these and all other MuSERA features, including interactive and batch execution as well as input/output standard data formats, is illustrated in the following sections; they show the relevance and broad utility of MuSERA for biological investigation, and in particular of the graphical displays of the computational results that MuSERA provides in support of the biological interpretation of the performed NGS experiments.

# 2 MuSERA Interactive Mode

MuSERA runs in two modes, **Interactive Mode** and **Batch Mode**. The **Interactive Mode** integrates features to study a variety of characteristics for the input samples and their combined replicates. MuSERA interface in the **Interactive Mode** consists of four main sections as follows (see Figure 1):

1. Switch between **Interactive Mode** and **Batch Mode**
2. **Cached Samples box**, to add new sample(s), and display cached sample(s) and feature(s)
3. **Sessions box**, to create new sessions, show defined sessions, and save results
4. **Details box**, shows the details of the selected sample/feature or session.

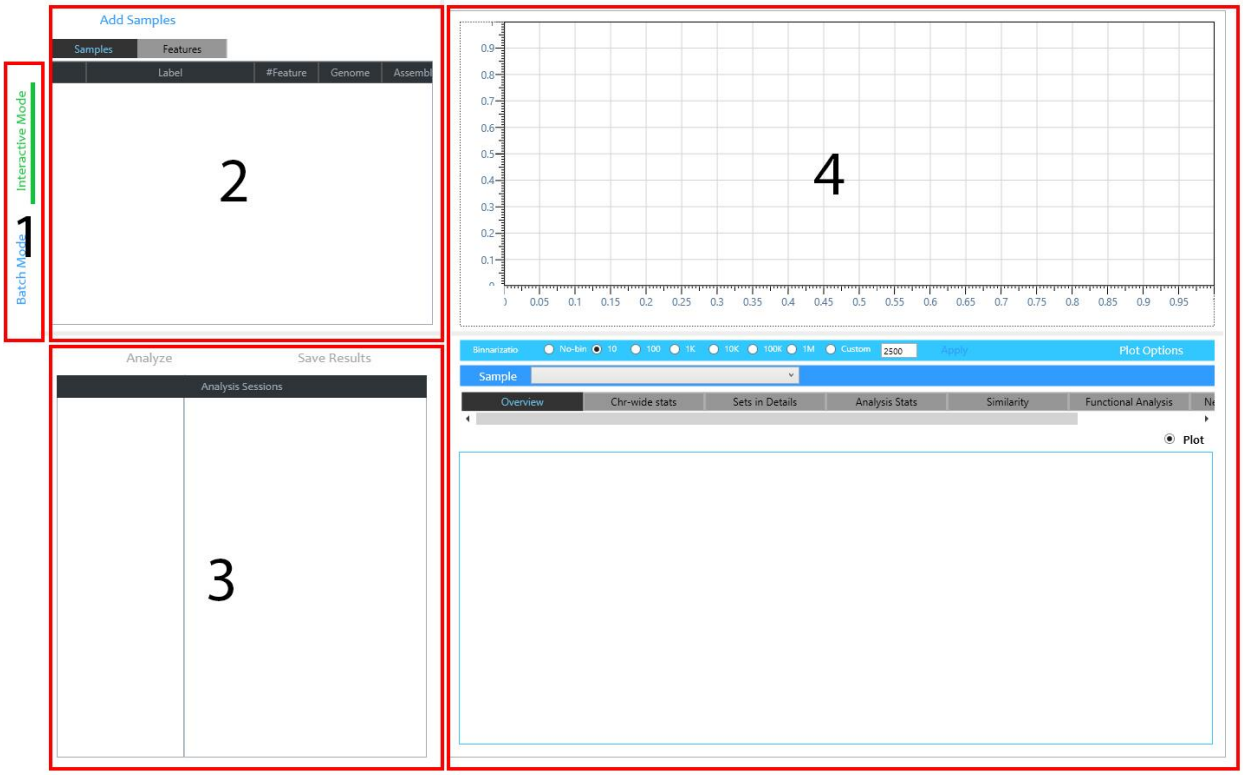


*Figure 1. MuSERA interactive mode interface.*

## 2.1 Cached Samples

### 2.1.1 Add Samples

Before it is possible to process samples, these need to be loaded and cached by the program. The caching procedure parsers the samples into in-memory data structures for real-time performance when operations are performed on them. To add a new sample, click on **Add Samples** button; a new window opens (see Figure 2) which consists of three sections as follows:

1. Browse: to open the file browser window to select input files.
2. Parser parameters: to setup the parsing parameters as it best matches the input file.
3. Add / Cancel: to add selected input file samples by invoking parser / to close the opened window.

To add a sample:

1. Click on the **Browse** button; in the window browser go to the folder containing your file(s); choose the appropriate file type from the drop-down menu on the bottom-right corner, then select your file(s) and click on **Open** button.
   Note: You may select multiple homogeneous files together. For instance, you may select multiple hg19 files with the same order of columns together. If you need to load different files (e.g., files with a different column order), you need to repeat the **Add Samples** procedure once for each.

2. Provide parser parameters (see section Parser Parameters) if required
3. Click the **Add** button.

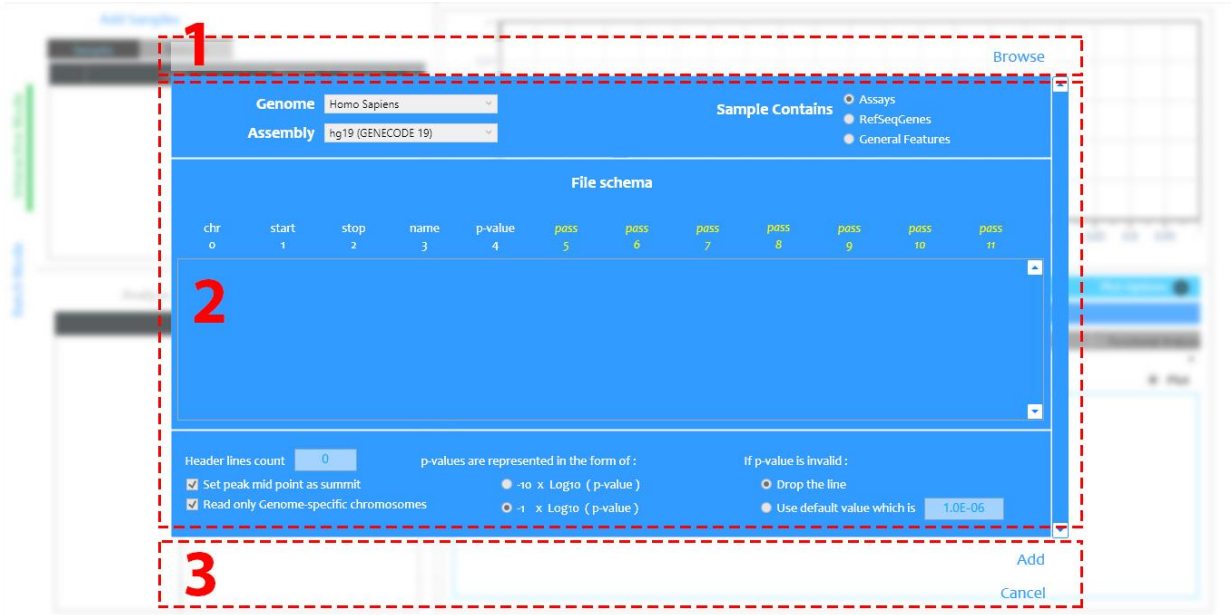


*Figure 2. Load a new sample interface*

### 2.1.2 Parser Parameters

You may provide parameters in the following order:

1. Specify the genome of the sample(s) from the **Genome** drop-down menu
2. Choose the appropriate assembly from the **Assembly** drop-down menu[1]
3. Specify the type of the selected sample(s) by choosing appropriate option on the top-right corner, among **Assays**, **RefSeqGenes**, or **General Features**. Doing so, MuSERA sets the parser column order to default values according to the selected content type.
4. Adjust the **File Schema** if needed. Based on your selected content type, MuSERA targets default columns for required information. To ease this procedure, MuSERA loads a small portion

[1] The choice of **Genome** and **Assembly** helps MuSERA in providing (1) overview statistics of the samples that depend on the genome size, such as the genome coverage by the sample ERs, (2) improves the parsing of MuSERA by reading only genome specific chromosomes and avoid ERs with chromosome labels such as *RAND_chr1* which are possible outputs of some peak callers (e.g., MACS [3]), and (3) tags the samples with the genome and assembly selection to help user in selecting the samples of related species and assemblies for the analysis. MuSERA currently supports *Homo sapiens* (with hg19 / GENCODE 19 assembly) and *Mus musculus* (with mm9 / GENCODE M2 assembly) for **Genome** and **Assembly** selection. However, the selected option has no impact on any of the analysis features MuSERA provides. To process samples of genomes/assemblies other than hg19 and mm9, user can de-select the "**Read only Genome-specific chromosomes**" option. Note that, the coverage of ERs is still calculated based on the genome size of the selected species (and assembly), and also on the cached samples box (see section 2 on Figure 1) the selected genome and assembly are reported for the sample, however, these information should be disregarded.

of your selected file (only from the first file if multiple files have been selected) and displays such portion as a table with color-assistant as follows:

- If the column content matches the expected value type, it will be displayed with a regular background, i.e., the color schema of the MuSERA application (see Figure 3)
- If the column content does not match the expected value type, it will be highlighted in red (see Column 3 in the table of Figure 3)
- If the column content contains acceptable value types, but these are not exactly as expected, they will be highlighted in yellow (e.g., a column containing numerical values can be used as column of the name of enriched regions; however, it may not be the appropriate selection since a sequence of alphabetic characters is expected. In such situation, MuSERA will highlight the selected column in yellow) (see Column 4 in the table of Figure 3)

In case of nonstandard formats, to match the required column to the correct column of your file, you can move the *cursor* over the required column name; this highlights a dark-blue box with pointers to left and right. Clicking on these pointers swaps the parser columns. For instance, in Figure 3 the parser is expecting for p-value on Column 3, while the p-value is provided by Column 4. Additionally, since the value type of the fourth column of the file is not double/integer (as required for p-value) the column is highlighted in red. To match the appropriate columns, the *cursor* can be placed on the **p-value** tag and, by clicking on the pointer to right, the **name** and **p-value** column can be swapped.

5. Your file may have multiple header lines; if so, you need to specify their count on the left-bottom corner of the Parser Parameters window. If the count parameter is not set, or the value is less than the number of header lines of the file, the parser will ignore the lines whenever they do not comply the column value type requirements.

When **Assays** is selected:

6. MuSERA requires the summit of the ERs (i.e., the position within each ER**,** i.e**.,** peak**,** with maximum accumulation of fragments determined by peak callers). If you check **set peak midpoint as summit** (on the left-bottom corner of the window), MuSERA uses the ER midpoint as summit; otherwise, the column of the file representing summit shall be set.
7. Some samples may contain ERs with chromosomes which are not defined in the selected Genome. MuSERA allows including or excluding such ERs through the **Read only Genome-specific chromosomes** option.
8. Different data sources provide p-values in logarithmic scale, and some other may multiply the logarithmic value by 10. To avoid the need for an external conversion, MuSERA can parse the p-value based on the selected conversion option (i.e., **-1xLog$_{10}$ (p-value)** or **-10xLog$_{10}$ (p-value)**) which is displayed on the center bottom part of the window.

9. Some data sources provide ERs without a p-value under specific conditions. MuSERA allows discarding such ERs, or assigning a user-defined p-value for all such ERs (using radio buttons on the bottom-right corner of the window).


*Figure 3. GUI features for loading samples.*

## 2.2 Sample Tables

MuSERA groups loaded samples in two separated tables as **Samples** containing assays (i.e., the samples parsed as **Assays** – see point 3 on 2.1.2**)** and **Features** including RefSeq genes and general features (i.e., samples parsed as **RefSeq genes**, or **General Features –** see point 3 on 2.1.2) (see Section 2 of Figure 1). Note that if no genomic features were loaded, the second table (**Features**) remains empty. The tables provide minimal information of the parsed samples as follows:

### 2.2.1 Samples table:

- **Sample color**: is a color assigned at random to each sample which will be used to differentiate each sample from the others when multiple ones are in a single plot. The color is user-modifiable (by double click on the default color and choosing a new color using the color selection tool).
- **Label**: by default is the file name of the sample (including the file extension); it is used when MuSERA refers to the sample. This label is user-modifiable (click on the label and type-in new label - you may need to press **F2** button to enable rename).
- **Feature count**, is the number of ERs extracted from the sample. This number is read-only.
- **Genome**, is the Genome specified when parsing the sample. This value is read-only.
- **Assembly**, is the assembly specified when parsing the sample. This value is read-only.

### 2.2.2 Feature table:

- **In use**: multiple sources of features might be loaded to MuSERA (see Section 2.1.1); however, the user may choose specific ones to be used at a time, based on the selected samples (e.g., when analyzing hg19 samples, features of the same assembly are used while other assemblies might be loaded as well)**.** Tick the checkbox for the desired features (a checkbox is available for each of the loaded feature sources).
- **Label**, by default is the file name of the feature (plus file extension) which will be used when MuSERA refers to the feature set; the label is user-modifiable (click on the label and type a new label, you may need to press **F2** button to enable rename).
- **Feature count,** is the number of determined features of the sample. This number is read-only.
- **Genome,** is the Genome of the features specified when parsing/adding them. This value is read-only.
- **Assembly,** is the assembly of the features specified when parsing/adding them. This value is read-only.
- **Data Type**, specifies whether the features are RefSeqGenes or general features.

## 2.3 Sessions

The session section (Section 3 on Figure 1) consists of three parts, described as follows (Figure 4):

1. **Analyze** and **Save Results** buttons initiate an analysis and export the analysis results, respectively; while analyzing, a timer displays the elapsed time of the analysis (see **Error! Reference source not found.**).
2. A table showing all the defined analysis sessions, with default labels. The labels are user-modifiable (click on the label and type a new label, you may need to press **F2** button to enable rename).
3. A brief overview of the selected session (displayed by single click on the session label from session's table).

By double-clicking on a finished session, session details are provided (see 1.6.2 Session details)

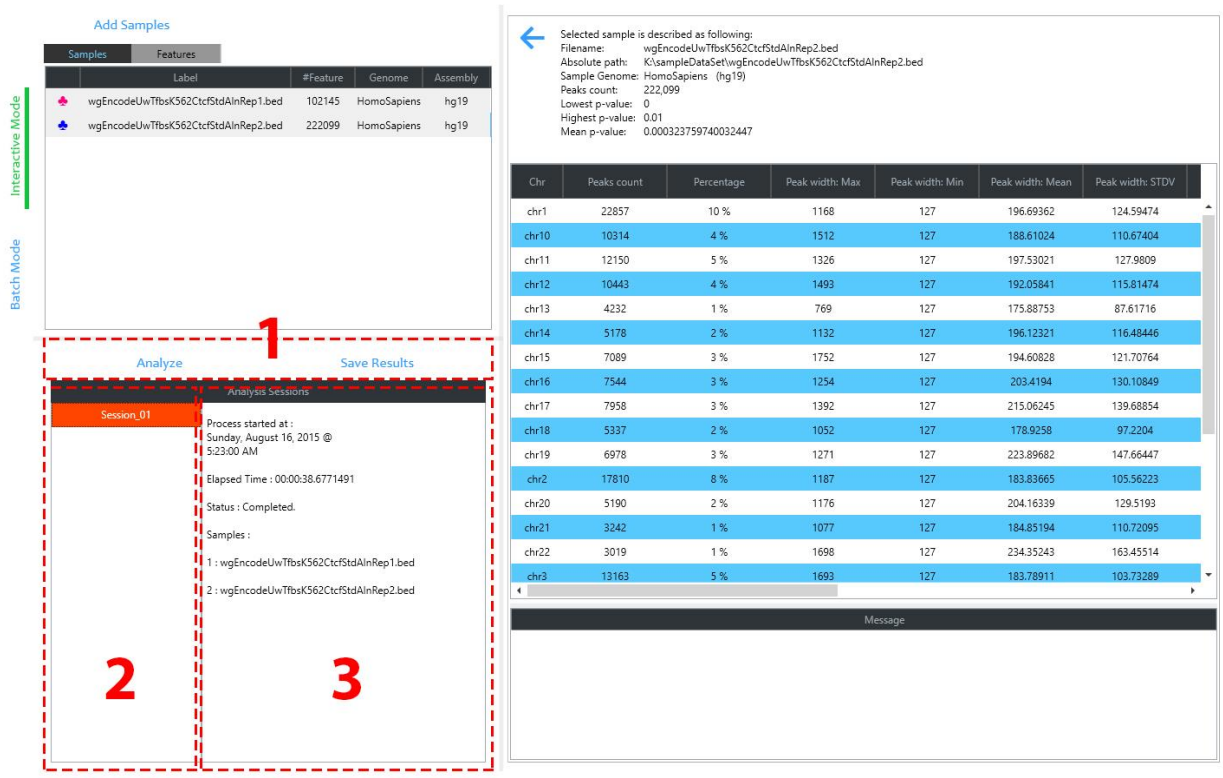


*Figure 4. Session panel*

## 2.4 Analyze Samples

To analyze samples:

1. Select the desired samples from the sample table. At least two samples are required for MuSERA to combine evidences.
2. Having selected the samples, the **Analyze** button in the session section (see section 3 of Figure 1) is enabled. Click on **Analyze** button.
3. Specify the desired analysis parameters in the pop-up window (see section Analysis Parameters).
4. Click on **Run Analysis** button in the pop-up window.
5. The pop-up window closes and the analysis starts.
6. While the analysis is being executed:
   - A timer displaying elapsed time is shown.
   - The analysis is executed by independent threads from the UI thread; therefore, while the analysis is ongoing, you may still use MuSERA for reviewing details of cached samples, reviewing previous sessions (if any), adding new samples, or continuing with **Batch mode**.
   - You are not allowed to start another analysis.
   - You are not allowed to save sessions.

### 2.4.1 Analysis Parameters

Analysis parameters are required for MuSERA to combine evidence from selected samples. The parameters are set through the pop-up window resulting by clicking on the **Analyze** button in Section 1 of Figure 4, and they are the following (Figure 5):

- **Replicate Type**, MuSERA treats biological and technical replicates differently; therefore, a replicate type has to be specified through the **Biological replicates** and **Technical replicates** radio buttons.
- **$T^s$**, specifies the stringency threshold (i.e., all ERs with p-values below this threshold are considered stringent evidence). The text box accepts only double values between 0 and 1.
- **$T^w$,** specifies the weak threshold (i.e., all ERs with p-values below this threshold, and above or equal $T^s$, are considered weak evidence). The text box accepts only double values between 0 and 1.
- $\boldsymbol{\gamma}$, specifies the combined stringency threshold. The text box accepts only double values between 0 and 1.
- $\boldsymbol{\chi^2_{\gamma,2k}}$ **,** is the $\chi^2$ of the specified $\gamma$ with $2k$ degrees of freedom, with $k$ being the number of selected samples. This text box is automatically updated when $\gamma$ text box is changed.
- **C**, sets the minimum number of overlapping ERs required to combine the p-values.
- **Multiple ERs overlapping**, it may happen that an ER from one replicate overlaps multiple ERs from another replicate. MuSERA considers only one overlapping ER per sample, therefore, through the **Use lowest p-value** and **Use highest p-value** radio buttons, MuSERA allows the user to choose between the most stringent or the least stringent among multiple overlapping ERs from the same replicate.
- **Multiple testing correction**, MuSERA corrects the p-values of the output ERs (the ERs passing tests, called "confirmed ERs" – for details refer to [1]) using the Benjamini-Hochberg multiple testing correction with user-specified false discovery rate ($\alpha$).

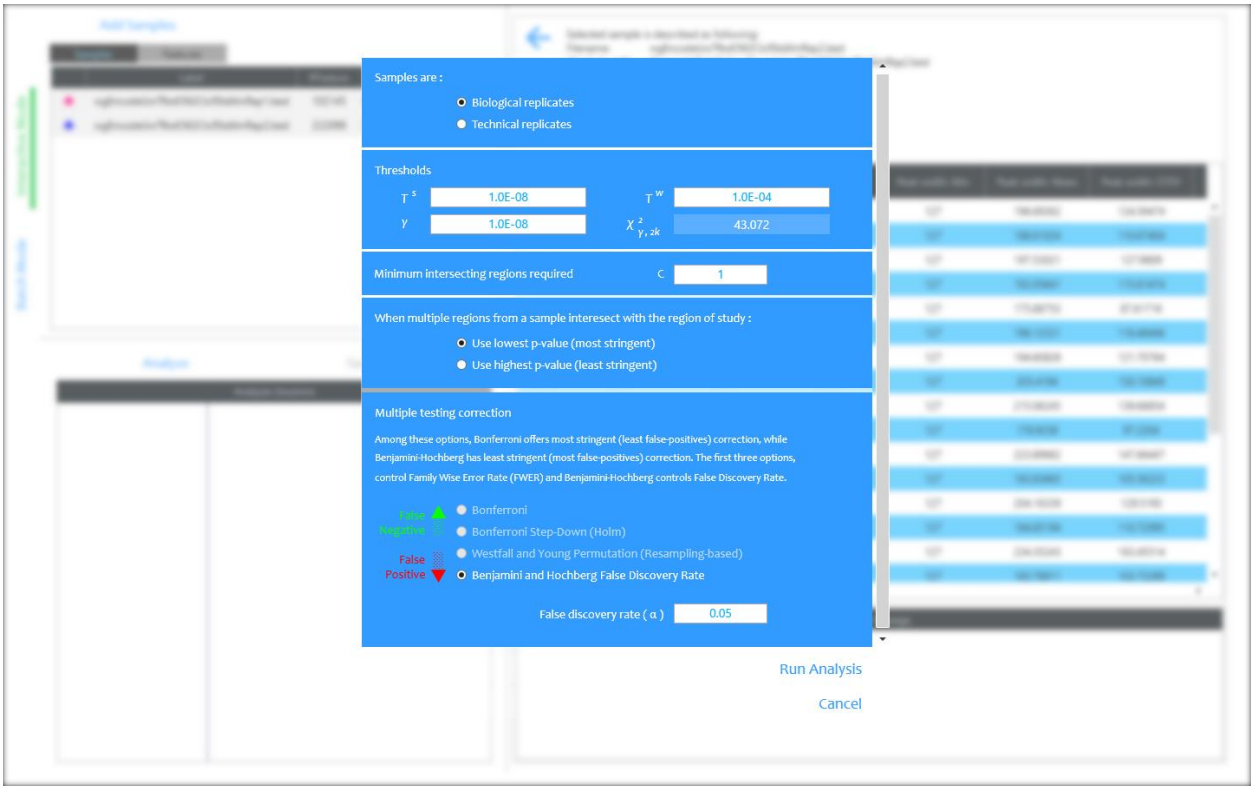


*Figure 5. The analysis parameter assignment interface.*

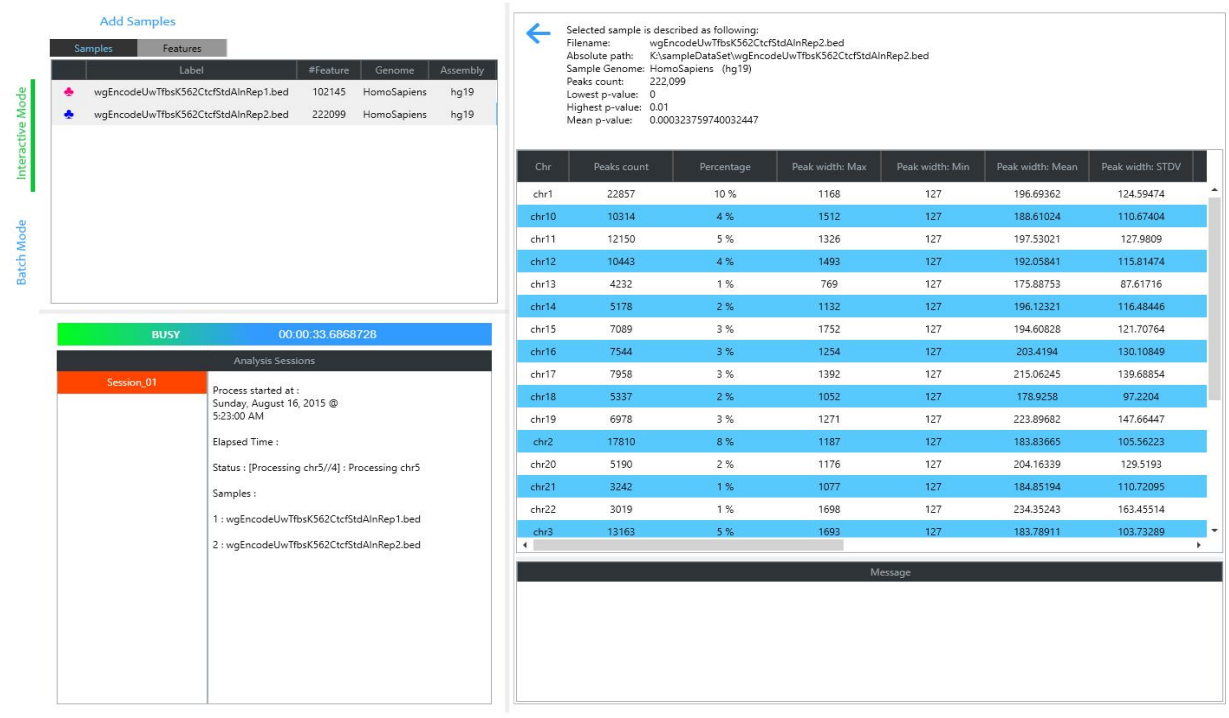


*Figure 6. Interactions while analysis is being performed. The analysis process is ongoing; meanwhile, a timer showing the elapsed time is displayed. Additionaly, the status of the analysis execution is given at selected session preview panel.*

## 2.5 Save Results

When available (i.e., no analysis is being executed), click on **Save Results** button (Section 1 on Figure 4) to export the available sessions using the save window (see Figure 7). To save:

1. Use the **Browse** button to specify the path for saving sessions.
2. From the table, check the **Save** column check box of the sessions you wish to export. The folder column is user-modifiable (by clicking on the label and entering a new label, and pressing **F2** if required), you may change it to any valid folder name for the session.
3. Choose the sets to save and their different types (if available) among BED and XML format.
4. If BED files are preferred without header, uncheck the **Add a header to BED files** checkbox.
5. Click on the **Save** button.
6. Close the save window when done, by clicking on the **Cancel** button

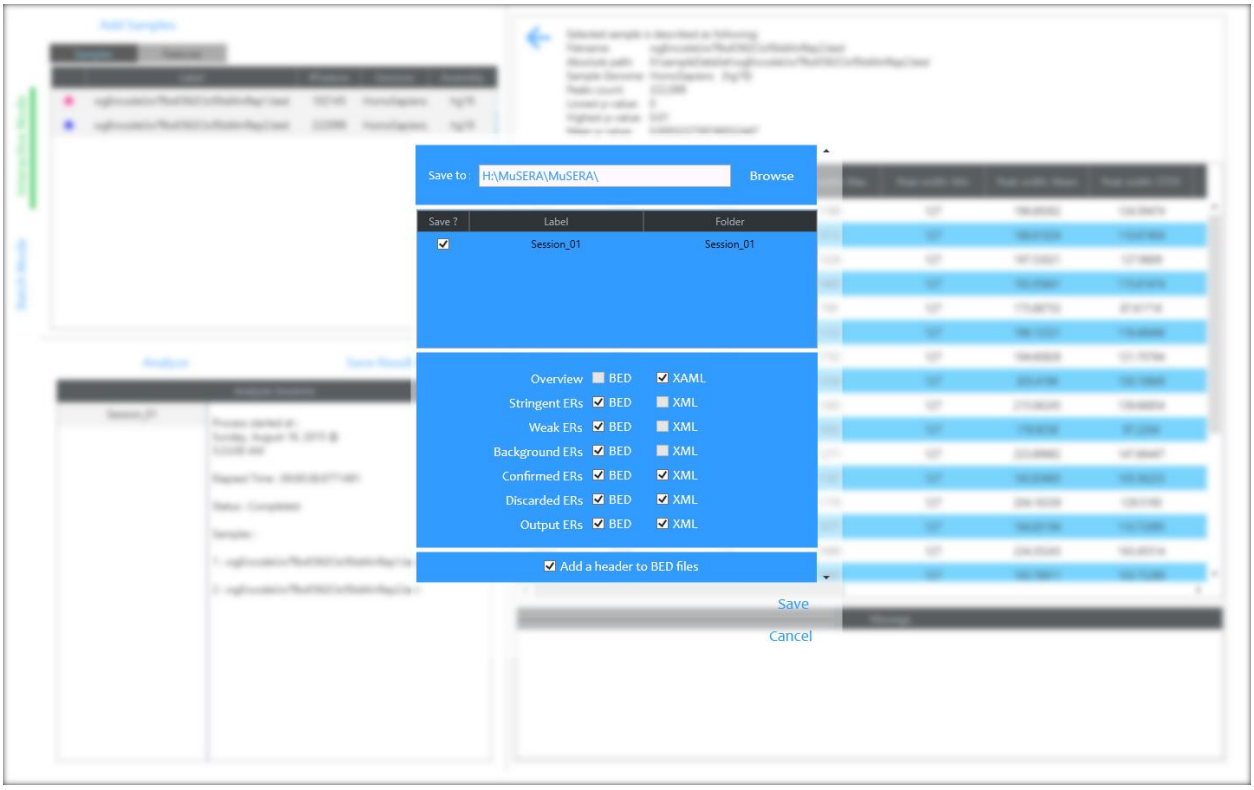


*Figure 7. The interface to save sessions.*

## 2.6 Details

The Detail Section (see Section 4 on Figure 1) provides detailed information for the selected cached data and the analysis session.

### 2.6.1 Cached data details

MuSERA provides details of the parsed samples/features, accessible by double-clicking on a sample/feature in the corresponding table. The provided information in the details section are: (i) a brief overview of the sample, (ii) chromosome-wide statistics, (iii) parsing messages (if any), such as the number of ignored lines and the reasons. See Figure 8.

By clicking on the blue back button (←), MuSERA displays the session details.

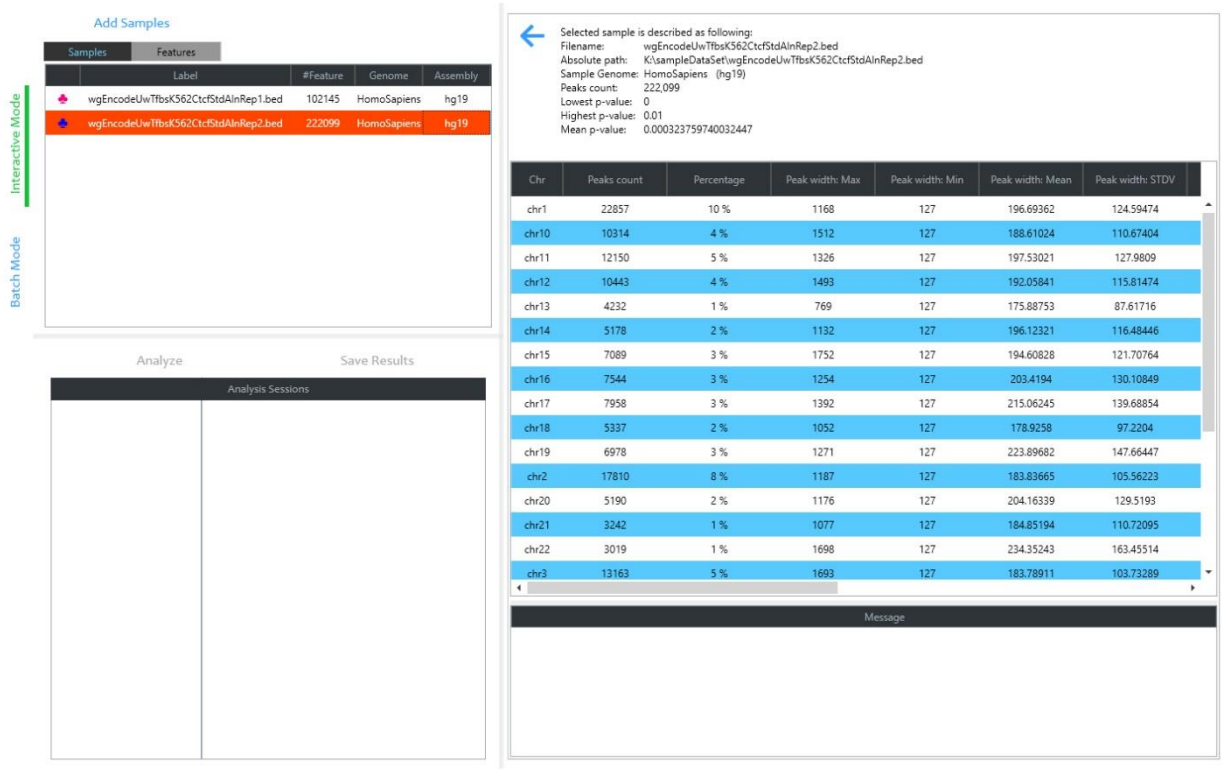


*Figure 8. Sample details panel.*

### 2.6.2 Session details

MuSERA provides comprehensive detailed information about the analysis, both at sample and session levels. Additionally, MuSERA integrates a wide range of common ER assessment features through a user-friendly GUI enabling further assessment of analysis results. Double click on a session label (from table in Section 2 of Figure 4) to load its details.

Note that some utilities require providing features (e.g., RefSeqGenes) and some others provide better insight (e.g., integrated genome browser) if features are given. The features in the **Features** table (see Figure 9) whose **In Use** check box is ticked are used for the double-clicked session. Once the session details are loaded, changes on the **In Use** check box have no effect on the selected features. For instance, if promoters of the genome assembly hg18 are ticked, after double-clicking on the session the hg18 promoters will be in use for the session features and changing the selection would not produce any change. To update the features selection (e.g., use hg19 promoters instead of hg18 promoters): change feature selection (e.g., uncheck hg18 promoters and check hg19 promoters) then double-click again on the session label.

The session details panel is structured in four sections, described as following (Figure 10):

1. An interactive plot, generated using Dynamic Data Display [2], a user-friendly library that allows:
   - Zooming: you may zoom on either one or both directions by scrolling up/down when the cursor is hovering on horizontal/vertical axis, or the plotting area, respectively.
   - Panning: you may change the section of the data to be displayed in the plot. This feature is particularly useful to display different genomic sections.
   - Saving snapshots: you may save the plot you are viewing in a PNG/JPEG/BMP/GIF file format (right click on the plot to access a drop-down menu).
2. Plotting options, which allow:
   - Bin size specification, setting MuSERA to bin the data based on the provided bin size (options are: **no-bin, 10, 100, 1K, 10K, 100K, 1M**, and a **Custom** size in base-pair)
   - **Plot Options** configuration, opening a window which allows specification of a few plotting options such as enable/disable legend, data label font size, etc.
3. Sample selection, allows to choose the sample of the session from the drop-down component whose details are required. When sample selection is changed, all tables and plots update automatically accordingly.
4. A tab-controlled panel, where each tab provides information of a specific aspect of the selected session and sample. The tabs are described in following sections.

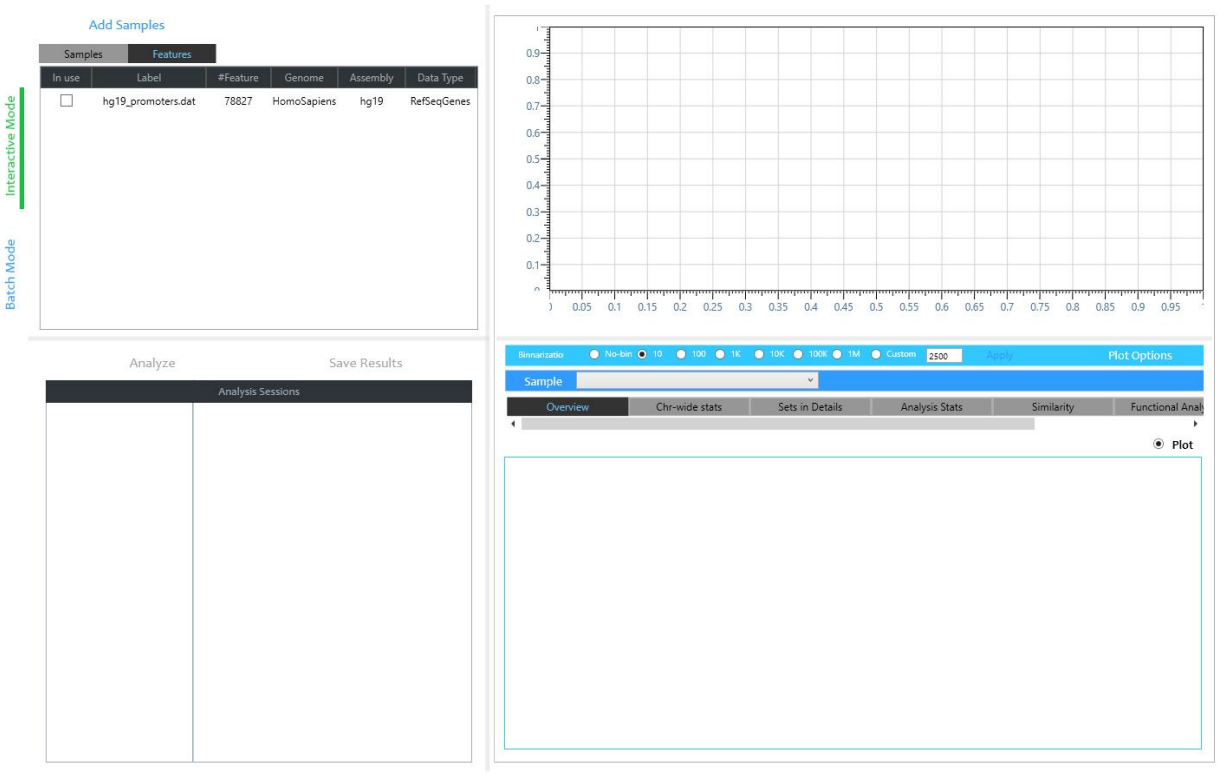


*Figure 9. Features tab*

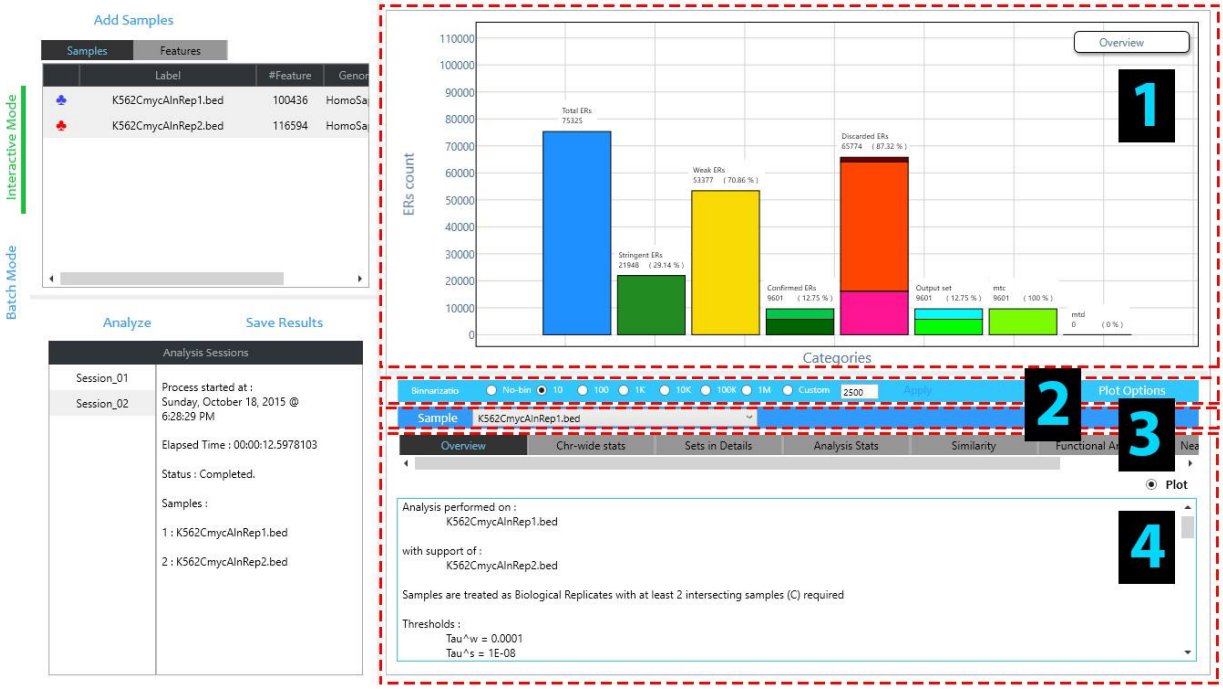


*Figure 10. Session details panel.*

### 2.6.3 Overview tab

The tab provides a detailed overview of the selected sample and summarizes all analysis parameters. Additionally, it provides detailed information of the selected sample ERs distribution in different sets as in Table 1. When the **Plot** radio button (on the top-right corner of the panel) is selected, this information is plotted as in Section 1 of Figure 10.

Note that the values of the cardinality, ratio, accumulated cardinality and accumulated ratio of intermediate sets of the analysis (all sets excluding input, output and multiple testing corrected ones) can possibly be higher than expected. For instance, if an ER fails one test and passes another, it will appear in both confirmed and discarded sets, and therefore $|R_j^c| + |R_j^d| > |R_j|$ when $|R_j^b| = \emptyset$. In other words, the set $R_j^c \cap R_j^d$ is not necessarily empty. A similar situation may happen with other sets (excluding input, output and multiple testing corrected ones).

### 2.6.4 Chr-wide stats tab

This tab provides chromosome-wide detailed information of the distribution of ERs in different sets (along with the corresponding ratios). The information are presented in a table and plotted, if the **Plot** radio button (on the top-right corner of the panel) is selected. See Figure 11. The provided information is the one reported in Table 2 **for each chromosome**:

| Count | Percentage | Description |
|---|---|---|
| $T = \lvert R_j / R_j^b \rvert$ | | Total number of ERs of the sample (excluding background ERs) |
| $\lvert R_j^s \rvert$ | $\lvert R_j^s \rvert / T$ | Stringent ERs set |
| $\lvert R_j^w \rvert$ | $\lvert R_j^w \rvert / T$ | Weak ERs set |
| $\lvert R_j^b \rvert$ | $\lvert R_j^b \rvert / \lvert R_j \rvert$ | Background ERs set |
| $\lvert R_j^c \rvert$ | $\lvert R_j^c \rvert / T$ | Weak and stringent confirmed ERs set |
| $\lvert R_j^d \rvert$ | $\lvert R_j^d \rvert / T$ | Weak and stringent discarded ERs set |
| $\lvert R_j^o \rvert$ | $\lvert R_j^o \rvert / T$ | ERs in output set |
| $\lvert R_j^{mtc} \rvert$ | $\lvert R_j^{mtc} \rvert / \lvert R_j^o \rvert$ | Multiple-testing confirmed ERs. |
| $\lvert R_j^{mtd} \rvert$ | $\lvert R_j^{mtd} \rvert / \lvert R_j^o \rvert$ | Multiple-testing discarded ERs. |
| $\lvert R_j^{sc} \rvert$ | $\lvert R_j^{sc} \rvert / \lvert R_j^s \rvert$ | Stringent confirmed ERs set |
| $\lvert R_j^{sdc} \rvert$ | $\lvert R_j^{sdc} \rvert / \lvert R_j^s \rvert$ | Stringent ERs discarded for failing to comply minimum overlapping ERs requirement. |
| $\lvert R_j^{sdt} \rvert$ | $\lvert R_j^{sdt} \rvert / \lvert R_j^s \rvert$ | Stringent ERs discarded for failing to comply minimum combined stringency requirement. |
| $\lvert R_j^{wc} \rvert$ | $\lvert R_j^{wc} \rvert / \lvert R_j^w \rvert$ | Weak confirmed ERs set |
| $\lvert R_j^{wdc} \rvert$ | $\lvert R_j^{wdc} \rvert / \lvert R_j^w \rvert$ | Weak ERs discarded for failing to comply minimum overlapping ERs requirement. |
| $\lvert R_j^{wdt} \rvert$ | $\lvert R_j^{wdt} \rvert / \lvert R_j^w \rvert$ | Weak ERs discarded for failing to comply minimum combined stringency requirement. |
| | $\lvert R_j^{sc} \rvert / \lvert R_j^o \rvert$ | * Cardinality of stringent-confirmed ERs set divided by the cardinality output set |
| | $\lvert R_j^{wc} \rvert / \lvert R_j^o \rvert$ | * Cardinality of weak-confirmed ERs set divided by the cardinality output set |
| | $\lvert R_j^{sco} \rvert / \lvert R_j^o \rvert$ | + Cardinality of stringent-confirmed ERs set which are in output set divided by the cardinality output set. |
| | $\lvert R_j^{wco} \rvert / \lvert R_j^o \rvert$ | + Cardinality of weak-confirmed ERs set which are in output set divided by the cardinality output set |
| | $\lvert R_j^{sc} \rvert / T$ | Stringent confirmed ERs set |
| | $\lvert R_j^{sdc} \rvert / T$ | Stringent ERs discarded for failing to comply minimum overlapping ERs requirement. |
| | $\lvert R_j^{sdt} \rvert / T$ | Stringent ERs discarded for failing to comply minimum combined stringency requirement. |
| | $\lvert R_j^{wc} \rvert / T$ | Weak confirmed ERs set |
| | $\lvert R_j^{wdc} \rvert / T$ | Weak ERs discarded for failing to comply minimum overlapping ERs requirement. |
| | $\lvert R_j^{wdt} \rvert / T$ | Weak ERs discarded for failing to comply minimum combined stringency requirement. |

*Table 1. The statistical information overview of the analyzed samples. The list contains information on the count and ratio of different ER sets, this information is provided by the Overview tab on session details (**Interactive Mode**) and Log file (**Batch Mode**). The two ratios marked with + and * might differ depending on the analysis parameters (e.g., in case of technical replicate, if an ER is confirmed in one test and fails another, then the ER will not be available in output set).*

| Provided by | | | |
|---|---|---|---|
| Table | Plot | Information | Description |
| ✓ | ✓ | $\left\|R_j\right\|$ | Total number of ERs in of the chromosome |
| ✓ | ✓ | $\left\|R_j^s\right\|$ | Total number of stringent ERs of the chromosome |
| ✓ | | $\left\|R_j^s\right\| / \left\|R_j\right\|$ | Ratio of stringent ERs of the chromosome |
| ✓ | ✓ | $\left\|R_j^w\right\|$ | Total number of weak ERs of the chromosome |
| ✓ | | $\left\|R_j^w\right\| / \left\|R_j\right\|$ | Ratio of weak ERs of the chromosome |
| ✓ | ✓ | $\left\|R_j^c\right\|$ | Total number of confirmed ERs of the chromosome |
| ✓ | | $\left\|R_j^c\right\| / \left\|R_j\right\|$ | Ratio of confirmed ERs of the chromosome |
| ✓ | ✓ | $\left\|R_j^d\right\|$ | Total number of discarded ERs of the chromosome |
| ✓ | | $\left\|R_j^d\right\| / \left\|R_j\right\|$ | Ratio of discarded ERs of the chromosome |
| ✓ | ✓ | $\left\|R_j^o\right\|$ | Total number of ERs in output set of the chromosome |
| ✓ | | $\left\|R_j^o\right\| / \left\|R_j\right\|$ | Ratio of ERs in the output set of the chromosome |
| ✓ | ✓ | $\left\|R_j^{mtc}\right\|$ | Total number of multiple-testing confirmed ERs of the chromosome |
| ✓ | | $\left\|R_j^{mtc}\right\| / \left\|R_j^o\right\|$ | Ratio of multiple-testing confirmed ERs of the chromosome |
| ✓ | ✓ | $\left\|R_j^{mtd}\right\|$ | Total number of multiple-testing discarded ERs of the chromosome |
| ✓ | | $\left\|R_j^{mtd}\right\| / \left\|R_j^o\right\|$ | Ratio of multiple-testing discarded ERs of the chromosome |

*Table 2. The list of information provided chromosome-wide.*

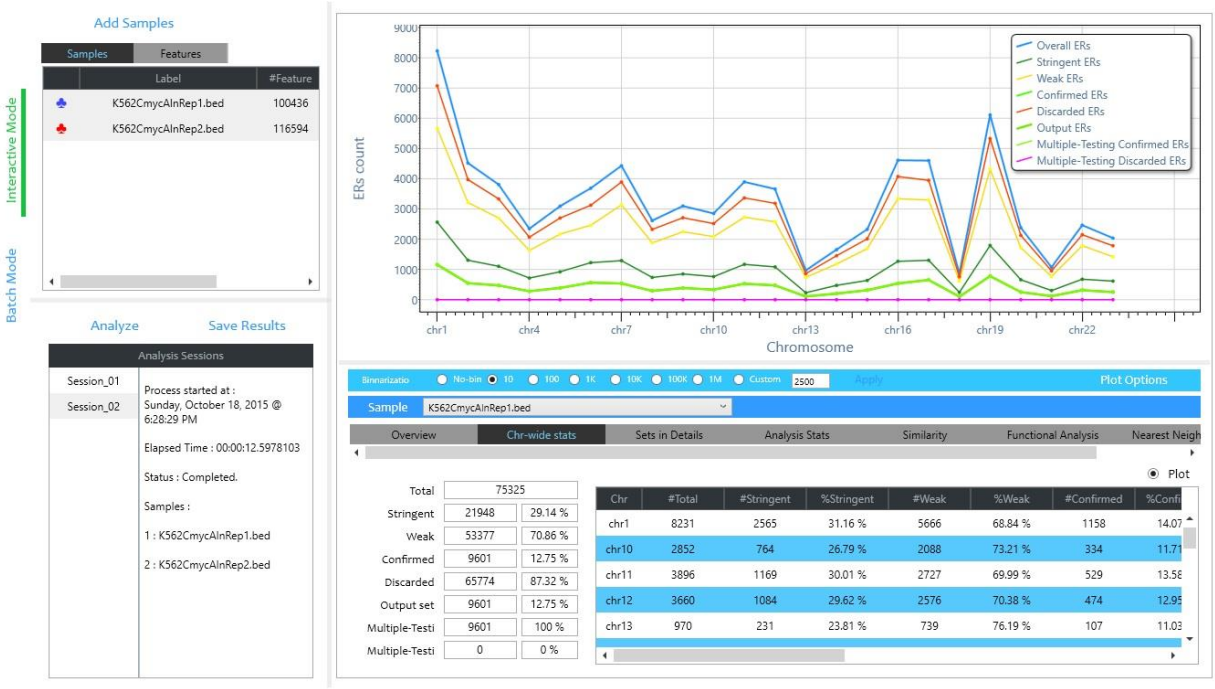


*Figure 11. Chromosome-wide distribution of enriched regions in different sets, displayed both via a table and plot.*

### 2.6.5 Sets in details tab and Integrated Genome Browser

This tab shows all the ER sets of the selected samples, along with available information for each ER in tables. Additionally, an **Integrated Genome Browser** can plot a selected ER and all its neighbors on the same sample, or different samples, and the provided feature set. The integrated genome browser visualizes the neighboring genomic regions of the selected ER by allowing zooming in/out and panning. By hovering the cursor on the left/right-end of an ER, a tooltip displays a brief explanatory information of the ER (see Figure 12). The genome browser maintains the marker size (on left and right ends) of ERs, so that even at large scales (considerable zooming-out) the accumulation of ERs on genomic regions can be highlighted.

The tab consists of five different sections described as follows (Figure 12), which shall be followed in this order:

1. Sets are grouped in chromosomes: to review sets, a desired chromosome has to be selected from the drop-down component in Section 1 (Figure 12).
2. To review a set, choose the appropriate tab from Section 2 (Figure 12). Each tab provides a table containing all ERs for the selected set and chromosome with all available information for each ER.

| Column / Set | Stringent ERs | Weak ERs | Confirmed ERs | Discarded ERs | Output ERs | Background ERs |
|---|---|---|---|---|---|---|
| Chromosome | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Left-end | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Right-end | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Summit | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Name | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| p-value | ✓ | ✓ | ✓ | ✓ | ✓ | ✓ |
| Classification (e.g., stringent-confirmed) | | | ✓ | ✓ | ✓ | |
| $X^2$ | | | ✓ | ✓ | ✓ | |
| Right-tail probability | | | ✓ | ✓ | ✓ | |
| Supporting ERs | | | ✓ | ✓ | ✓ | |
| Reason of discarding | | | | ✓ | | |
| Adjusted p-value | | | | | ✓ | |

*Table 3. Information provided for enriched regions in each set.*

3. Choose the number of dichotomies you would like to see on the integrated genome browser. Dichotomies are contiguous portions on the genome determined by ERs of samples of the selected session, where ERs in a portion are intersecting and none of the ERs of any two portions intersect with each other (portions are disjoint). Note that, the plot updates accordingly when an ER is double-clicked. When the number is set to $n$, $n$ dichotomies surrounding the selected ER are displayed on the integrated genome browser.
   You may enable/disable plotting genes and general features on the integrated genome browser, if provided, by selecting the corresponding options in Section 3 (Figure 12).
4. When a new session is selected for detail visualization, MuSERA populates some data structures in background. A progress bar (shown in Section 4 of Figure 12) displays the current progress. In the meanwhile, you may use MuSERA for all features that do not depend on the completion of this process. For plotting a region (and its dichotomies) on the integrated genome browser, we highly recommend to wait for the completion of this process, although MuSERA will not prevent plotting (because the information you are demanding might be ready), but the provided information might be incomplete.
5. Section 5 (Figure 12) contains a table for each ER set. You may review the information for any ER in the corresponding table; you may refer to Table 3 for the list of provided information for each ER in the different sets. Additionally, you may double-click on an ER to visualize it on the integrated genome browser (e.g., Figure 12; on this figure we have renamed the labels of the samples to *rep1*, *rep2*, and *promoters* for clearer understanding). The browser will zoom to the ER and all supporting ERs (if available) by default; you may zoom-out or pan to review

neighbors. To see details of ERs in the plot, you may hover the cursor over the left-end or right-end of the ER, which then displays a tooltip with short explanatory information about the ER.

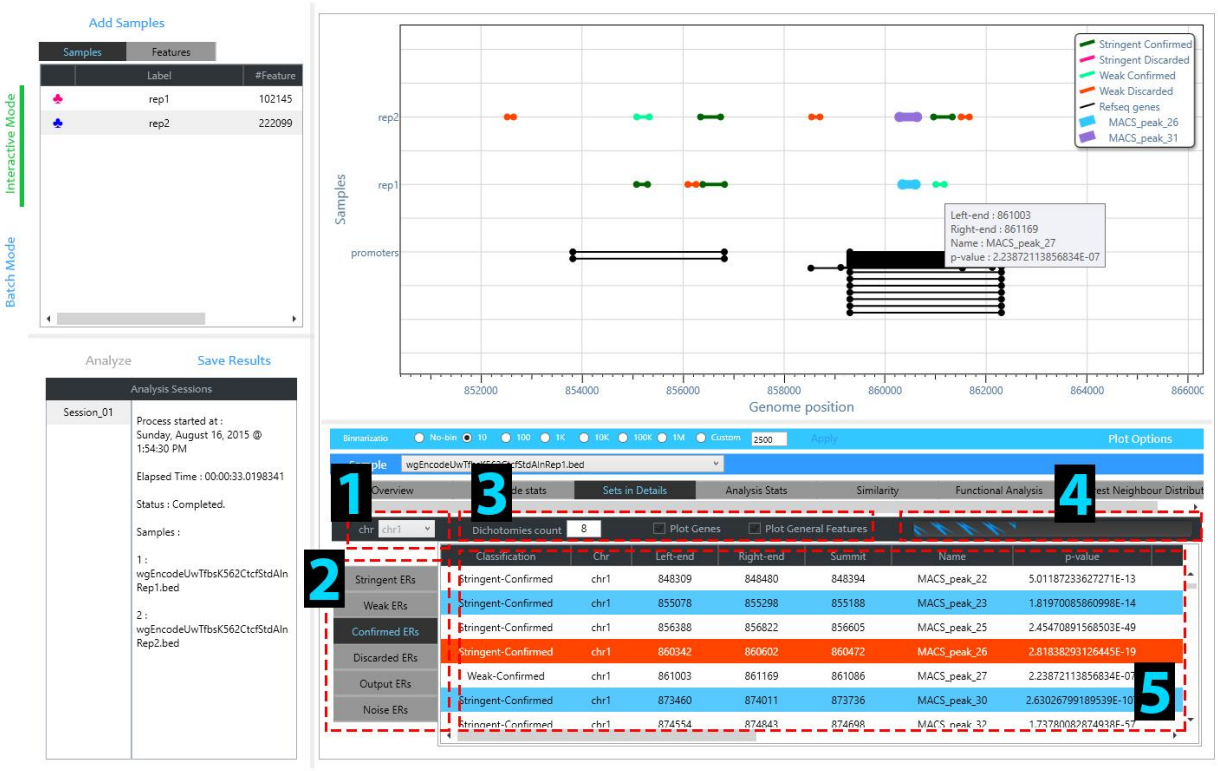


*Figure 12. The GUI of ER sets in details tab and Integrated Genome Browser.*

### 2.6.6 Analysis stats tab

The tab provides information about the distribution of the ER p-values using different tables and plots (see Figure 13). For better assessment, the distributions are provided for different groups as follows:

- Input set, including background ERs – see Section 1 on Figure 13, providing the distribution on all the samples of selected session.
- Combined p-values (see Section 2 on Figure 13)
- 1st Classification: Input ERs are classified as stringent or weak (or background, which is excluded from this tab stats). See Section 3 on Figure 13.
- 2nd Classification: stringent and weak ERs are assessed using replicates, and they are classified as confirmed or discarded. See Section 4 on Figure 13.
- 3rd Classification: confirmed and discarded sets are further processed and an output set is created, which consists of stringent and weak confirmed ERs. See Section 5 on Figure 13.
- 4th Classification: a multiple testing correction procedure is executed on ERs in output set, which provides multiple-testing confirmed or discarded ERs. See Section 6 on Figure 13.

Note that although combined p-values are included in 2nd classification, a separated table enables a more clear visualization of combined stringency. You may change the bin size of the plotting using available binning options, highlighted at Section 2 on Figure 13.

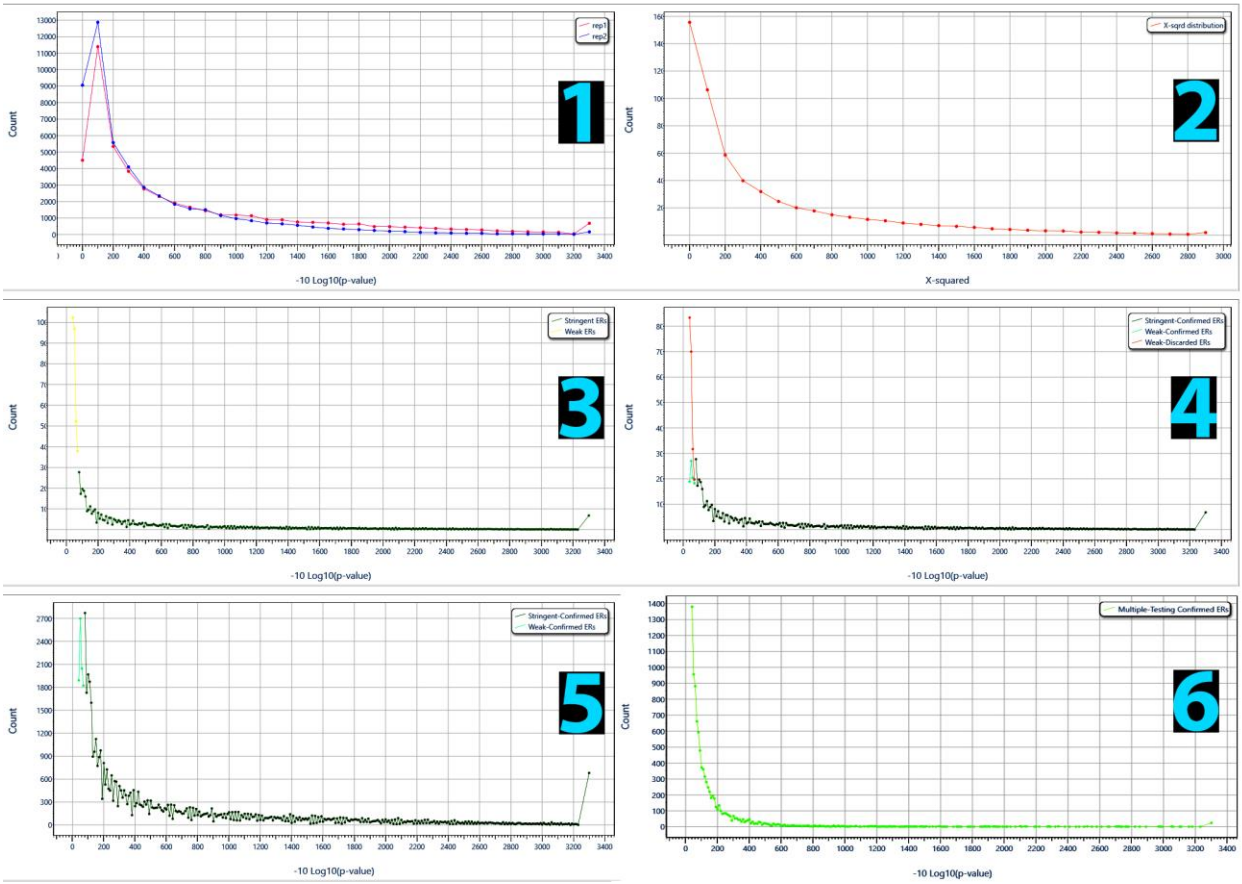


*Figure 13. Distributions of p-values in different classifications. The plot provides only available information; for instance, since no ER is classified as stringent-discarded, no plot line is plotted on the section **4** plot. Similarly, the multiple testing discarded set is not provided in section **6** plot.*

### 2.6.7 Similarity tab

This tab provides similarity between samples of the selected session from a variety of perspectives. The similarities are provided in terms of both **base-pair-level** and **region-level** Jaccard similarity index **J**. MuSERA provides:

- **Overall** similarity: computed across samples in terms of regions and base-pairs in common (Figure 14); the overall similarity between two replicates in Figure 14 is computed as follows:
  **Base-pair-level:**

$$J = \frac{|Rep1 \cap Rep2|}{|Rep1 \cup Rep2|}$$

$$= \frac{(6-4)+(19-16)+(27-24)+(49-46)+(58-56)}{(10-2)+(30-13)+(37-33)+(43-40)+(51-46)+(60-54)} = 0.3$$

**Region-level:**

$$J = \frac{|Rep1 \cap Rep2|}{|Rep1 \cup Rep2|} = \frac{4+5}{6+5} = 0.81$$

- **Sample-wide** similarity: we compute similarities as follows:
  - How similar is Rep1 to Rep2 in Figure 14?

    **Base-pair-level:**

$$J = \frac{\sum|r_i|: r_i \in Rep1 \ \wedge \ r_i \cap Rep2 \ \neq \emptyset}{\sum|r_i|: r_i \in Rep1}$$

$$= \frac{(6-4)+(19-16)+(27-24)+(49-46)+(58-56)}{(6-4)+(19-13)+(27-24)+(51-46)+(58-54)} = 0.65$$

    **Region-level:**

$$J = \frac{|\{r_i | r_i \in Rep1 \ \wedge \ r_i \cap Rep2 \ \neq \emptyset\}|}{|Rep1|} = \frac{5}{5} = 1$$

  - How similar is Rep2 to Rep1 in Figure 14?

    **Base-pair-level:**

$$J = \frac{\sum|r_i|: r_i \in Rep2 \ \wedge \ r_i \cap Rep1 \ \neq \emptyset}{\sum|r_i|: r_i \in Rep2}$$

$$= \frac{(6-4)+(19-16)+(27-24)+(49-46)+(58-56)}{(10-2)+(30-16)+(37-33)+(43-40)+(49-46)+(60-56)} = 0.3$$

    **Region-level:**

$$J = \frac{|\{r_i | r_i \in Rep2 \ \wedge \ r_i \cap Rep1 \ \neq \emptyset\}|}{|Rep2|} = \frac{4}{6} = 0.6$$

The similarity indexes are plotted in Figure 15; as illustrated, for the sample-wide similarities, the similarity index of Rep1-to-Rep2 is higher than the index of Rep2-to-Rep1. These similarity indexes indicate that all Rep1 ERs are intersecting with ERs of Rep2 (region-level similarity of Rep1 to Rep2 = 1; base-pair-level similarity of Rep1 to Rep2 = 0.65: 65% of base pairs of Rep1 ERs are overlapping with Rep2 base pairs), while Rep2 has ERs not overlapping Rep1 ERs (region-level similarity of Rep2 to Rep1 = 0.6: only 60% of Rep2 ERs are overlapping Rep1 ERs; base-pair-level similarity of Rep2-to-Rep1 = 0.3

indicates that only 30% of the base-pairs of Rep2 ERs are covered by Rep1 ERs). The plot also highlights the differences between different similarity assessments.

MuSERA provides these similarity assessments between different sets of samples. Information is provided in a table where rows are grouped by **overall**, and **sample-wide** sections. The similarity estimates help to better understand the different ER sets. For instance, a sample may contain a large number of ERs when peaks are called with a permissive p-value threshold, which are not necessarily co-localized with evidences on other replicates. Therefore, similarity between input sets is expected to be low; the similarity between discarded evidences is expected to be very low, while higher similarities are expected for confirmed ERs and output sets. A hierarchical representation of different classes of ERs along with their similarity indexes is provided in Figure 16 (MuSERA provides the information required to generate this figure, the figure is not generated by MuSERA itself).

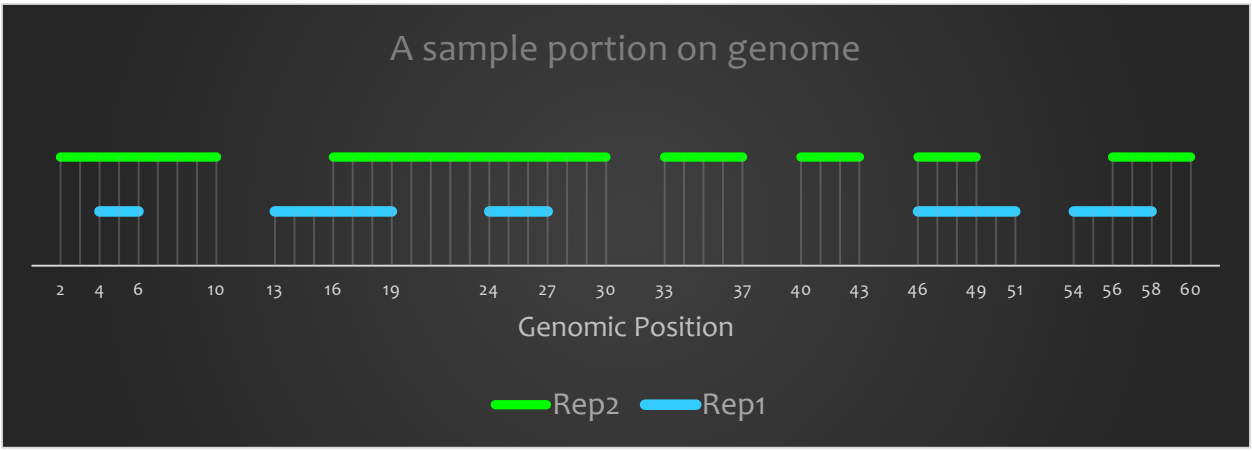


*Figure 14. An example of a portion of genome with two replicates.*

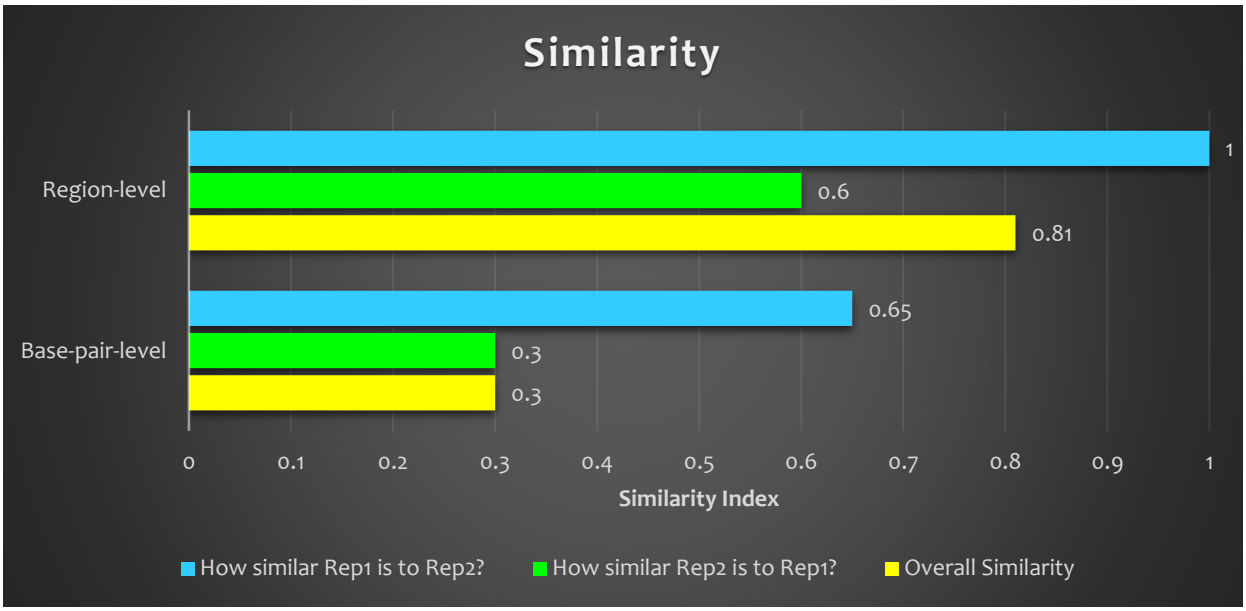


*Figure 15. The estimated similarities of the example discussed in the text.*

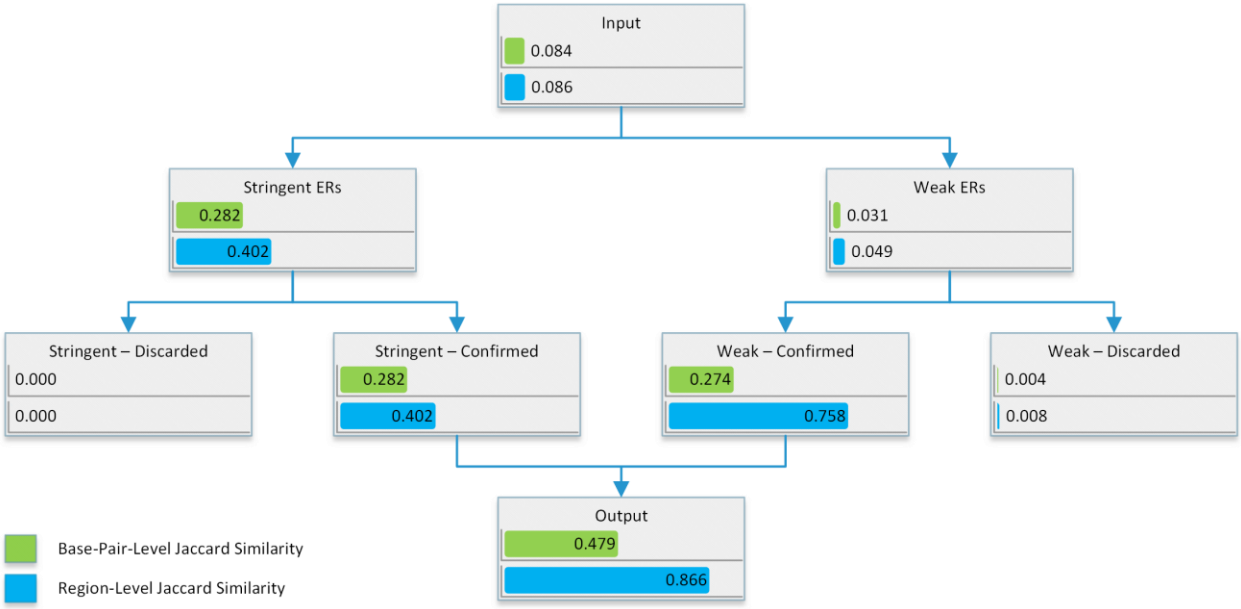


*Figure 16. The hierarchy of enriched region classification and their Jaccard similarity indexes.*

### 2.6.8 Functional analysis tab

This tab helps assessing the distances between ERs and given genomic features (see Figure 17). To do so, i.e., to calculate the distance from set $X$ to set $Y$, MuSERA requires the user to provide the sets $X$ and $Y$, where $X$ is a set of ERs and $Y$ is a set of genomic features. For instance, $X$ can be the set of all weak-

confirmed and weak-discarded ERs on chr2 and chrX of all the samples of the session; and $Y$ can be a selected source of genomic features.

To correctly perform the functional analysis, you may proceed as follows:

**Specifying $X$**

1. Click on the dropdown button of **Source Samples** and tick the checkbox of the samples you want to include. The unticked samples will not be considered in the process.
2. Click on the dropdown button of **Chr** and tick the checkbox of the chromosomes you want to include. The unticked chromosomes will not be considered in the process.
3. Click on the dropdown button of **ER Classification** and choose the classifications you want to include. Mind the hierarchy of classifications (see Figure 16); for instance, if you choose stringent, then both stringent-confirmed and stringent-discarded ERs will be included in the process.

**Specifying $Y$**

4. If the session is loaded with RefSeq genes dataset, then the **use Genes** radio button is enabled. If a general feature dataset is selected for the session, then the **use General Features** radio button is enabled. However, only one dataset at a time is allowed. You may choose your desired dataset.
5. Click on the **Update** button.

When the process is completed, the corresponding table will be populated with the proper distances and counts. Additionally, you may activate the **Plot** radio button to plot the distribution. You may also change the bin size of the distance distribution to obtain desired approximations.

The **distance** is the distance between the summit of an ER belonging to the $X$ selection and the closest feature belonging to the $Y$ selection.

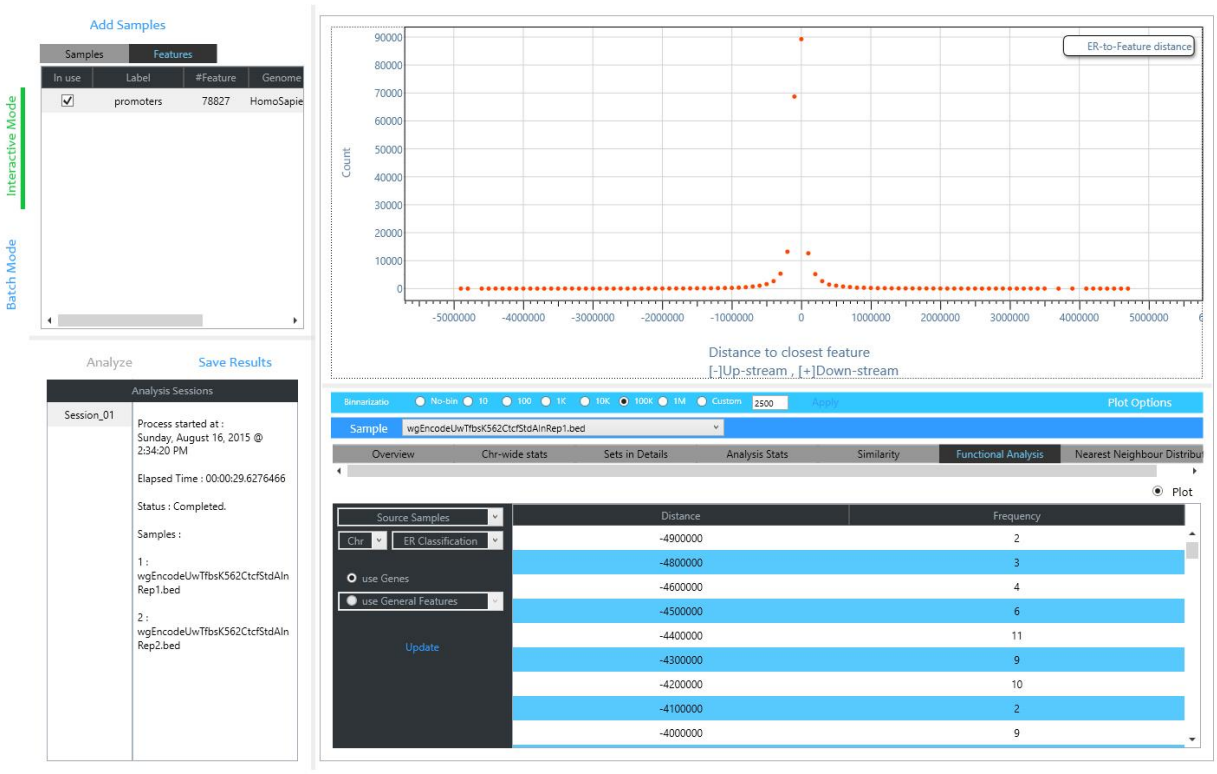


*Figure 17. The functional analysis panel.*

### 2.6.9 Nearest Neighbor Distribution tab

This tab provides features to estimate the distribution of distances between the ERs and their nearest neighbors (see Figure 18). You may proceed as follows to correctly use it:

1. Click on the dropdown button of **Source Samples** and tick the checkbox of the samples you want to include. The unticked samples will not be considered in the process.
2. Click on the dropdown button of **Chr** and tick the checkbox of the chromosomes you want to include. The unticked chromosomes will not be considered in the process.
3. Click on the dropdown button of **ER Classification** and choose the classifications you want to include. Mind the hierarchy of classifications (see Figure 16); for instance, if you choose stringent, then both stringent-confirmed and stringent-discarded ERs will be included in the process.
4. If you choose more than one classification, you may choose to combine or to keep separate their ER distributions using **Differentiate ER Classifications** checkbox. If ticked, it separates the ER distributions, and if not ticked it combines the ER distributions.

The **distance** is the distance between an ER belonging to the selection and the closest ER also belonging to the selection. For instance, let us suppose that weak-confirmed and weak-discarded classifications are chosen. For each ER belonging to the weak-confirmed or weak-discarded classification, the distance

to the closest weak-confirmed or weak-discarded ER will be determined. If the closest ER is upstream, the distance is between the start of the ER and the end of the neighbor. If the closest ER is downstream, the distance is between the end of the ER and the start of the neighbor. If the closest neighbor is overlapping the ER, then the distance is zero.

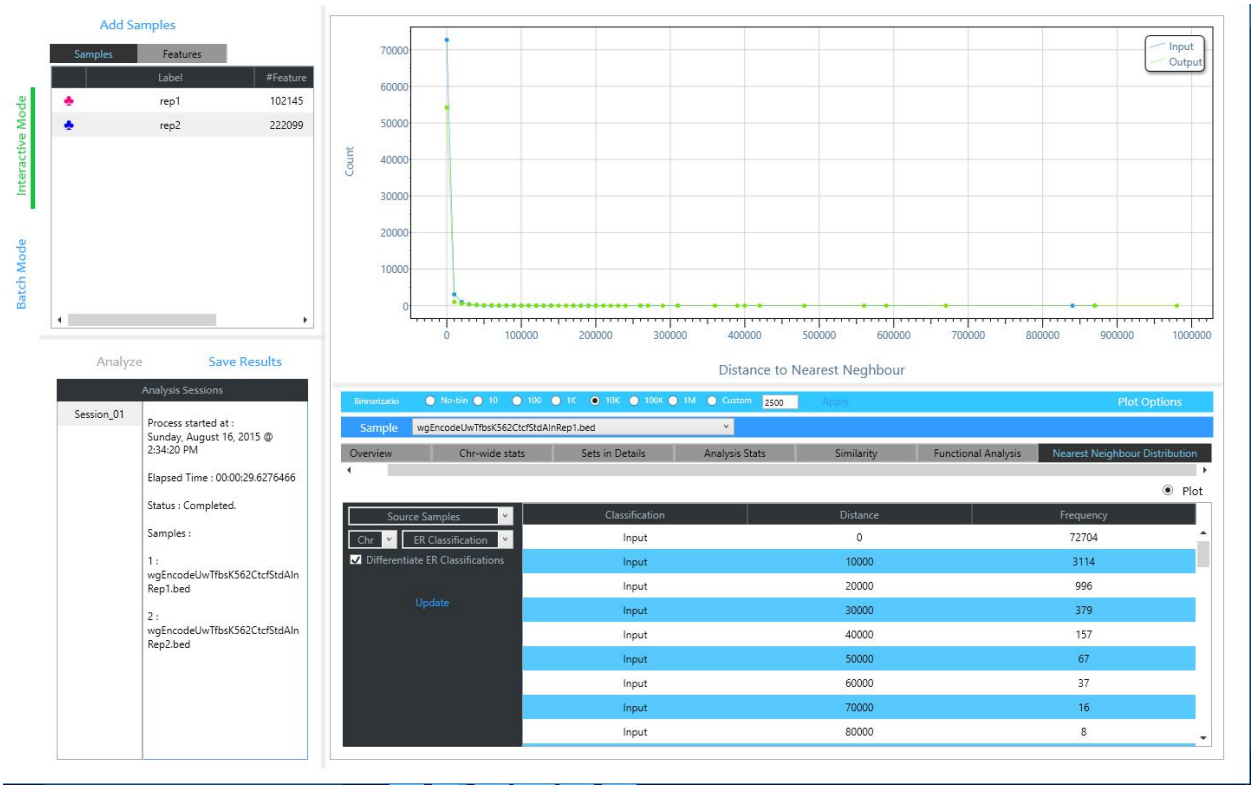


*Figure 18. The nearest neighbor distribution panel.*

# 3 MuSERA Batch Mode

The **Batch Mode** is intended to combine a large number of replicates through an automatic procedure with pre-defined parameters provided by a simple XML structure. In general, **Batch Mode** takes as input the XML file that defines an *at-Job*, executes the defined analysis, and outputs the results to the defined location; in addition, it provides statistics for all executed analysis. The input and output of **Batch Mode** are explained in the following subsections.

## 3.1 Input: at-Job

Let an *at-Job* be the collection of tasks to be executed by MuSERA in **Batch Mode**. The *at-Job* is defined in an XML file compliant with the World Wide Web Consortium (W3C) Document Object Model (DOM) Level 1 Core and the DOM Level 2 Core recommendations. The *at-Job* properties are set in the XML file as follows:

```
<Setter Property="PropertyName" Value="PropertyValue"/>
```

The ***PropertyName*** and ***PropertyValue*** are replaced by the name and value of the property to be set using this `Setter`. For instance:

```
<Setter Property="Alpha" Value="0.05" />
```

sets the $\alpha$ parameter of the analysis to $0.05$.

The XML file is mainly structured in three sections described in following subsections. Note that, MuSERA provides a template of *at-Job* file with all the parameters for two sample sessions to facilitate the definition of a custom *at-Job* file – see Section 3.2.1.

### 3.1.1 Plot parameters

This section provides the parameters for automatically generated plots (e.g., overview plot that is generated for each analyzed sample).

```
<Plot>
      <Setter Property="Width" Value="3500" />
      <Setter Property="Height" Value="1500" />
      <Setter Property="Font Size" Value="32" />
      <Setter Property="Axis Font Size" Value="32" />
      <Setter Property="Data Label Font Size" Value="32" />
      <Setter Property="Header font size" Value="40" />
      <Setter Property="Overview" Value="Enabled" />
</Plot>
```

The **`Width`** and **`Height`** properties set the resolution of the plot. The property **`Font Size`** sets the font size of the plot which is applied on plot legend and axis labels. The **`Axis Font Size`** property sets the font size of labels on the horizontal and vertical axes. The **`Data Label Font Size`** property sets the font size of data labels (i.e., the labels on markers of line charts or the text above bar plot). The property **`Header font size`** sets the font size of plot header. The property **`Overview`** enables and disables the overview plot automatically generated for each analyzed sample. All these parameters are optional and default values are given in Table 4.

| *Property* | *Default value* |
|---:|---|
| *Width* | 3500 |
| *Height* | 1500 |
| *Font size* | 32 |
| *Axis font size* | 32 |
| *Data label font size* | 32 |

| | |
|---|---|
| *Header font size* | 40 |
| *Overview* | Enabled |

*Table 4. Default plotting parameters*

### 3.1.2 Log File

After performing an analysis session, in addition to exporting the analysis results, MuSERA exports a set of statistics (see Table 1) for each analyzed sample; the information are provided using two tab-delimited files named *Log_TestIDs.txt* and *Log_TestStats.txt* that are saved into a folder defined as follows in the XML *at-Job* file:

```
<LogFile>
      <Setter Property="Path" Value="C:\Users\Vahid\Desktop\" />
</LogFile>
```

This key sets the folder of analysis statistics files (i.e., *Log_TestIDs.txt* and *Log_TestStats.txt*) to **C:\Users\Vahid\Desktop**. The *Log_TestStats.txt* file provides information as described in Table 1 for each of the analyzed samples. Each sample and the corresponding session (i.e., information on supporting samples and analysis parameters) are automatically assigned with an ID and the information on *Log_TestStats.txt* are identified using this ID. The sample and the session corresponding to the ID are given in *Log_TestIDs.txt* file, which has the structure explained in Table 5. The statistics given by the *Log_TestStats.txt* file provide a wide range of informative insights on the analyzed sample which reveals various aspects of the data; Figure 19 provides an example.

| Column number | information | Description | Example |
|---|---|---|---|
| *0* | Analysis Time | The time at which the analysis is performed | [2015-10-18 - 12:30:21 AM] |
| *1* | ID | The ID of the analysis | Session_01_1.0 |
| *2* | Replicate Type | It states if the sample is analyzed considering supporting samples being biological replicates or technical replicates. | Technical |
| *3* | $T^s$ | The stringency threshold | 1E-08 |
| *4* | $T^w$ | The weak threshold | 0.01 |
| *5* | γ | The combined stringency threshold | 1E-08 |
| *6* | Multiple testing correction | The choice of multiple testing correction | BenjaminiHochberg |
| *7* | α | The false-discovery rate of multiple testing correction | 0.05 |
| *8* | Analyzed sample | The sample that is analyzed | K562CmycAlnRep1.BED |
| *9* | Supporting samples | A semicolon (;) delimited list of supporting samples | K562CmycAlnRep2.BED; |

*Table 5. The structure of Log_TestIDs.txt file*

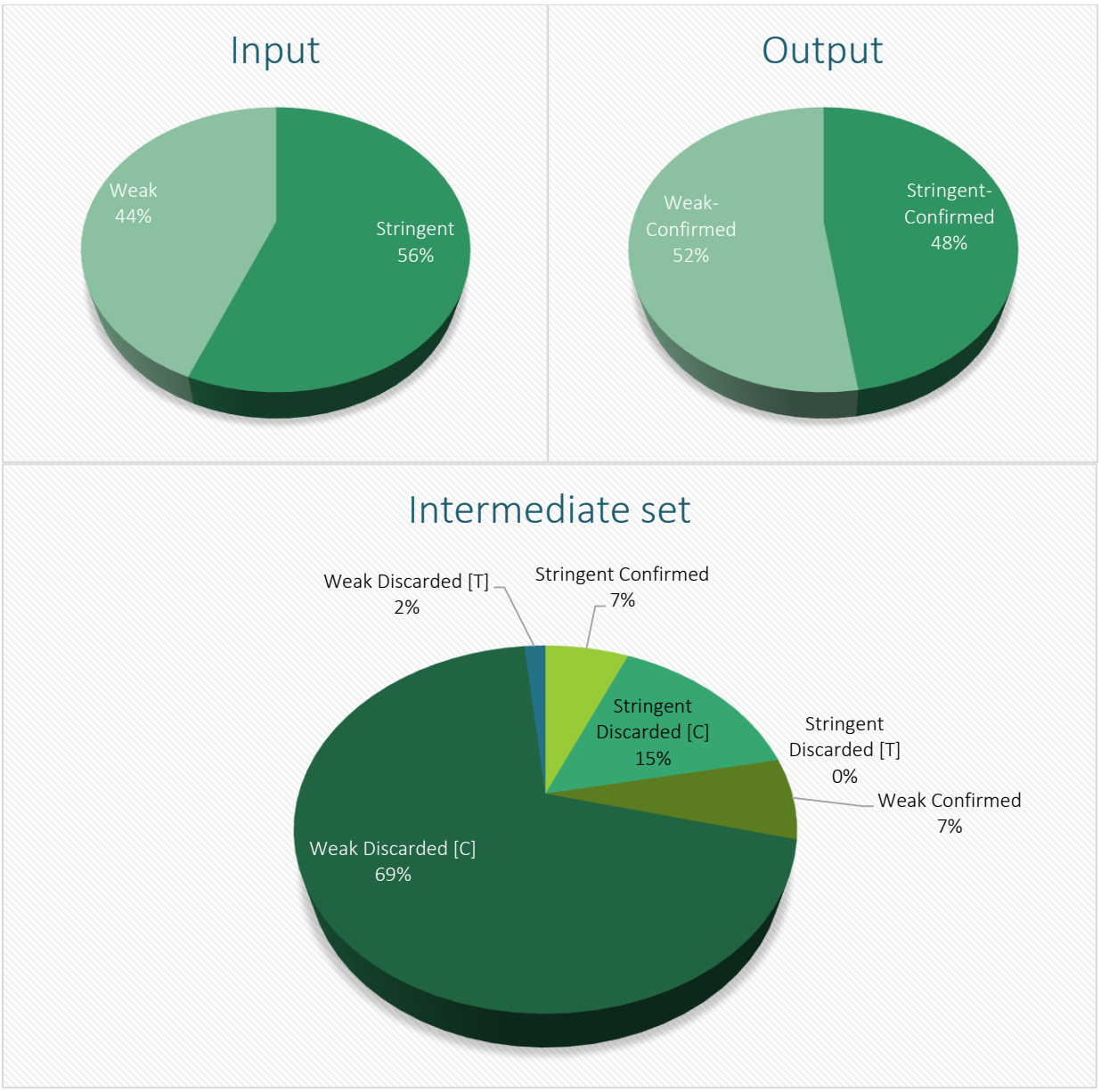


*Figure 19. Information provided by Log_TestStats.txt. The information for these pie-charts are a subset of the information provided by the Log_TestStats.txt file that are plotted here as an example. These are the information about sample* ***wgEncodeBroadHistoneK562CtcfStdAlnRep1****, analyzed with* ***wgEncodeBroadHistoneK562CtcfStdAlnRep2*** *as supporting sample, and considering them as biological samples with* $T^s$ ***= 1E-8***, $T^w$ ***= 0.01***, ***C = 2***, ***γ=1E-8***, *and* ***α = 0.05****. [C]: refers to ERs discarded due to failing the minimum overlapping ER requirement; and [T] refers to ERs discarded due to failing the combined stringency test.*

### 3.1.3 Sessions

A *session* is the analysis of two or more replicates that are combined using the *session parameters*. The main section of a **Batch Mode** *at-Job* file includes the sessions that define the analysis to be performed. An *at-Job* file may contain any number of sessions (at least one); sessions are processed independently,

and each has its own title that facilitates further reference. In general, a session is defined in the *at-Job* XML file as follows:

```
<Session Title="Session_01">
   <Load_and_Save_Parameters>
       <Setter Property="Input Sample" Value="D:\Data\K562CmycAlnRep1.BED" />
       <Setter Property="Input Sample" Value="D:\Data\K562CmycAlnRep2.BED" />
       <Setter Property="Output Path" Value="D:\Data\Results\" />
       <Setter Property="Export Output Set (BED)" Value="true" />
       <Setter Property="Export Output Set (XML)" Value="true" />
       <Setter Property="Export Stringent ERs" Value="true" />
       <Setter Property="Export Weak ERs" Value="true" />
       <Setter Property="Export Background ERs" Value="true" />
       <Setter Property="Export Confirmed ERs (BED)" Value="true" />
       <Setter Property="Export Confirmed ERs (XML)" Value="true" />
       <Setter Property="Export Discarded ERs (BED)" Value="true" />
       <Setter Property="Export Discarded ERs (XML)" Value="true" />
       <Setter Property="Export Chromosome-wide statistics" Value="true" />
   </Load_and_Save_Parameters>
   <Analysis_Parameters>
       <Setter Property="Replicate Type" Value="biological" />
       <Setter Property="TauS" Value="1.00E-8" />
       <Setter Property="TauW" Value="1.00E-2" />
       <Setter Property="Gamma" Value="1.00E-8" />
       <Setter Property="Alpha" Value="0.05" />
       <Setter Property="C" Value="2" />
       <Setter Property="Multiple Testing Correction" Value="BH FDR" />
   </Analysis_Parameters>
   <BED_Parser_Parameters>
       <Setter Property="Start Offset" Value="1" />
       <Setter Property="Chr Column" Value="0" />
       <Setter Property="Start Column" Value="1" />
       <Setter Property="Stop Column" Value="2" />
       <Setter Property="Name Column" Value="3" />
       <Setter Property="p-value Column" Value="4" />
       <Setter Property="Drop Line if no p-value" Value="true" />
       <Setter Property="Default p-value" Value="1.00E-6" />
       <Setter Property="p-value Conversion Option" Value="-1xLog10(p-value)"/>
   </BED_Parser_Parameters>
</Session>
```

A session has three sets of parameters as follows:

**A. Load and Save parameters** consists of the following properties:

**a. Input Sample:** Provides the *absolute path* (root directory, all sub-directories, and file name with extension) to the sample to be analyzed. An *at-Job* file contains one setter of this property for each sample (replicate). For instance:

```
<Setter Property="Input Sample" Value="D:\Data\K562CmycAlnRep1.BED" />
<Setter Property="Input Sample" Value="D:\Data\K562CmycAlnRep2.BED" />
```

defines two replicates to be analyzed.

**b. Output path:** is the absolute path of the directory to which the results of the analysis shall be saved. For instance:

```
<Setter Property="Output Path" Value="D:\Data\Results\" />
```

c. **Output set:** enables/disables saving output sets in BED and XML format. For instance:

```
<Setter Property="Export Output Set (BED)" Value="true" />
<Setter Property="Export Output Set (XML)" Value="true" />
```

d. **Stringent/Weak/Background ERs:** is a collection of properties that enable or disable saving the stringent, weak and background ERs to BED files; for instance:

```
<Setter Property="Export Stringent ERs" Value="true" />
<Setter Property="Export Weak ERs" Value="true" />
<Setter Property="Export Background ERs" Value="true" />
```

e. **Confirmed/Discarded ERs:** is a collection of four properties that enable or disable saving confirmed and discarded ERs to BED and XML files; for instance:

```
<Setter Property="Export Confirmed ERs (BED)" Value="true" />
<Setter Property="Export Confirmed ERs (XML)" Value="true" />
<Setter Property="Export Discarded ERs (BED)" Value="true" />
<Setter Property="Export Discarded ERs (XML)" Value="true" />
```

f. **Chromosome-wide statistics:** MuSERA determines chromosome-wide statistics for each of the analyzed samples (see Table 2), this property enables/disables saving these statistics to the output folder; for instance:

```
<Setter Property="Export Chromosome-wide statistics" Value="true" />
```

B. **Analysis parameters:** defines the parameters required to analyze the samples using the following properties:

a. **Replicate Type**; samples are either biological replicates, or technical replicates, and their processing varies accordingly. The property is set as follows:

```
<Setter Property="Replicate Type" Value="biological" />
```

b. **Thresholds**; the stringency ($T^s$), weak ($T^w$), combined-stringency ($\gamma$), and minimum overlapping ERs thresholds are defined using **TauS**, **TauW, Gamma, C** properties, respectively; for instance:

```
<Setter Property="TauS" Value="1.00E-8" />
<Setter Property="TauW" Value="1.00E-2" />
<Setter Property="Gamma" Value="1.00E-8" />
<Setter Property="C" Value="2" />
```

c. **Multiple testing correction:** MuSERA uses the Benjamini-Hochberg multiple testing correction with $\alpha$ false discovery rate, an example of this setter is as follows:

```
<Setter Property="Multiple Testing Correction" Value="BH FDR" />
```

C. **BED Parser parameters:** the samples are given by BED files with standard definition for the order of their columns; however, if the samples are not in standard column order, the parser of MuSERA is flexible and can adapt to such conditions. In this regard, each session contains a section for BED parser parameters that defines the properties to correctly parse the input samples. The properties of this section are as follows:

   a. **Header count;** a BED file commonly contains one line of header; however, a BED file may lack a heather line or may have multiple lines. The following property is used to specify the number of header lines:

```
<Setter Property="Start Offset" Value="1" />
```

   b. **Column order;** a collection of properties as follows, which define the column order:

```
<Setter Property="Chr Column" Value="0" />
<Setter Property="Start Column" Value="1" />
<Setter Property="Stop Column" Value="2" />
<Setter Property="Name Column" Value="3" />
<Setter Property="p-value Column" Value="4" />
```

   c. **ERs without p-value;** it is required that all the ERs in a sample have a p-value to process them with MuSERA. To provide maximum flexibility, MuSERA lets the user to decide how to treat ERs without a p-value. The user may decide to ignore all such ERs, or may decide to keep them, but assigning a default p-value to them. Such a decision is communicated to MuSERA using the following properties:

```
<Setter Property="Drop Line if no p-value" Value="true" />
<Setter Property="Default p-value" Value="1.00E-6" />
```

   d. **p-value conversion;** the p-value is commonly given according to two conversions, "$-\log_{10}$ p-value" or "$-10\log_{10}$ p-value". The BED parser of MuSERA is flexible on this point and allows user to specify the conversion used in his/her file without the need for performing the conversion beforehand using a third party tool. The conversion is specified using the following property:

```
<Setter Property="p-value Conversion Option" Value="-1xLog10(p-value)"/>
```

   The possible values are "-1xLog10(p-value)" or "-10xLog10(p-value)".

## 3.2 Input: GUI

MuSERA graphic user interface (GUI) in **Batch Mode** consists of four sections as follows:

- Browse, load, run and abort *at-job* (see Section 1 on Figure 20).
- The action to be taken when *at-job* processing is completed (see Section 2 on Figure 20).
- The priority of batch process (see Section 3 on Figure 20)

- The status of batch execution (see Section 4 on Figure 20).

The sections are described in following subsections.

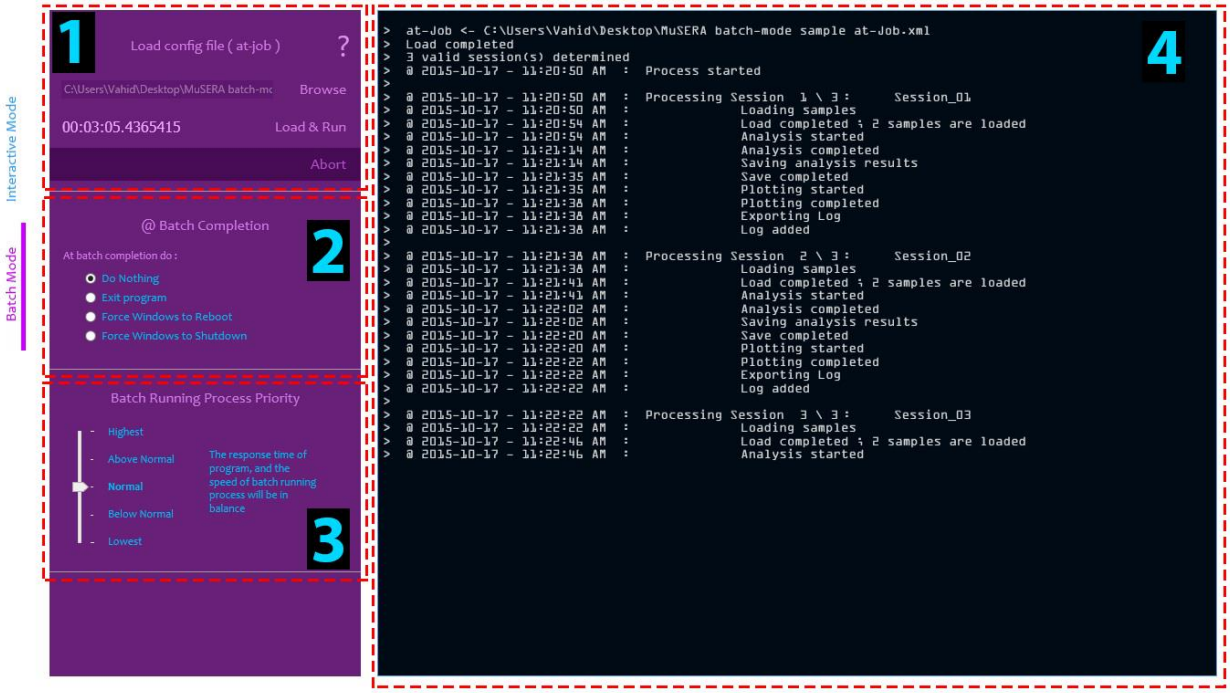


*Figure 20. Batch Mode GUI*

### 3.2.1 Load and Run Configuration file

To process a defined *at-Job* file in MuSERA, the following procedure is to be followed (with reference to Section 1 on Figure 20):

1. Click on **Browse** button and, in the browse window, select your *at-Job* file and click **Open** button.
2. Click on **Load & Run** button; at this point MuSERA parses the *at-Job* file and starts performing the defined analysis (the status can be tracked on Execution Status section)

The execution of the analysis defined by the *at-Job* file can be aborted at any time by clicking on **Abort** button.

MuSERA provides a sample *at-Job* XML file, which can be used as a template for defining any custom *at-Job* files. To obtain this file, click on the **?** label (upper-right corner on Section 1 of Figure 20) and then the file is created on the desktop of the used machine.

### 3.2.2 At Batch Completion

Depending on the number of the sessions defined in the *at-Job* file, and the number of ERs in the samples, the batch execution may take long time to complete. In such cases, MuSERA provides options of different actions to be taken at the completion of batch execution. The options are (with reference to section 2 on Figure 20):

a. **Do Nothing**: no action will be taken once the batch execution is completed.
b. **Exit Program**: exits MuSERA once the process is completed.
c. **Force Windows to Reboot**: reboots Microsoft Windows® machine, and asks the system to ignore any program or service preventing the reboot.
d. **Force Windows to Shutdown**: shuts down Microsoft Windows® machine and avoids any program or service preventing it.

### 3.2.3 Batch Priority

The **Batch Mode** can be used concurrently with **Interactive Mode**, or other programs on the machine might be running while the **Batch Mode** is busy executing the *at-Job* file. In such scenarios, MuSERA is flexible on modifying the priority of **Batch Mode** execution to best match the preferences on the scenarios. MuSERA provides five levels of priority for **Batch Mode** as: **Highest, Above Normal, Normal, Below Normal, and Lowest,** which can be set by simply adjusting the slider to the desired value (see Section 3 on Figure 20)**.** For instance, a user may prefer the batch execution to be completed faster while he/she is using also the **Interactive Mode** to perform another analysis, in this case he/she may set the **Batch Mode** priority to the **Highest** value.

### 3.2.4 Execution Status

MuSERA reports the execution status on a text box (see Section 4 on Figure 20). Any message, including status of execution, warnings and occurred errors are shown on this text box.

# References


1. Jalili V, Matteucci , Masseroli M, Morelli MJ. **Using combined evidence from replicates to evaluate ChIP-seq peaks.** *Bioinformatics.* 2015; **31**(17): p. 2761-2769.

2. Lyutsarev , Berezin , Microsoft research team. **Codeplex.** [Online].; 2015. Available from: http://dynamicdatadisplay.codeplex.com/wikipage?title=D3v1.

3. Yong Z, Liu , Meyer A, Eeckhoute J, Johnson S, Bernstein E, et al. **Model-based analysis of ChIP-Seq (MACS).** *Genome biology.* 2008; **9**(9): p. R137.