**MANUAL:**

**R Shiny App for Visualizing EndoC-βH1 and Human Islet Genomics Data**

**Author: Nathan Lawlor**

**Contact:** [**Nathan.lawlor03@gmail.com**](mailto:Nathan.lawlor03@gmail.com)

**Acknowledgements:**

**This application is a wrapper for the R package “Sushi”:**

**Douglas H Phanstiel (2015). Sushi: Tools for visualizing genomics data. R package version 1.16.0.**

**Sushi is available on Bioconductor at:** [**https://bioconductor.org/packages/release/bioc/html/Sushi.html**](https://bioconductor.org/packages/release/bioc/html/Sushi.html)

[**Interactive Tutorial**](#Interactive_Tutorial)

**Quick Start Guide:**

1. [**Select Data Types to Visualize**](#Select_Data_Types_to_Visualize)
   1. [EndoC-βH1](#a_EndoC_BH1)
   2. [Human Islet](#b_Human_Islet)
   3. [Color Schemes](#c_Color_schemes)
2. [**Specifying the Genomic Regions to View**](#Specifying_the_Genomic_Regions_to_View)
   1. [Search via gene symbol](#a_Search_via_gene_symbol)
   2. [Manually specify genomic coordinates](#b_Manually_specify_genomic_coordinates)
3. [**View the Data**](#View_the_Data)

[**Interacting with Tabbed Panels:**](#Interacting_with_Tabbed_Panels)

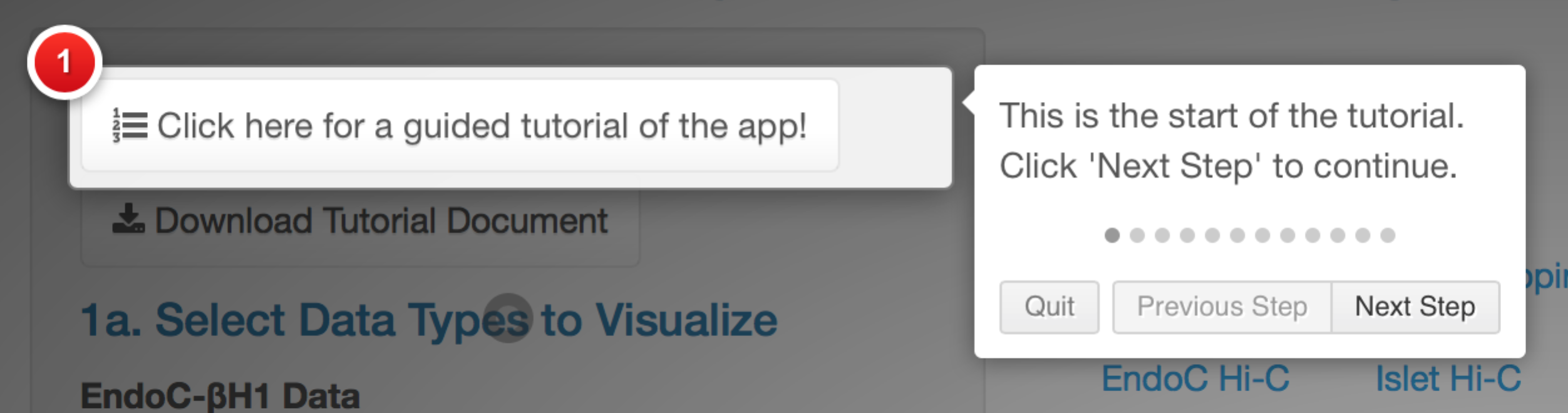
1. [**Multiomic Plot Panel**](#Multiomic_Plot_Panel)
2. [**Highlight Regions of Interest**](#Highlight_Regions_of_Interest)
3. [**Zoom In and Zoom Out Functions**](#Zoom_In_and_Zoom_Out_Functions)
4. [**Selected Coordinates**](#Selected_Coordinates)
5. [**Overlapping GWAS SNPs**](#Overlapping_GWAS_SNPs)
6. [**Nearby Chromatin Interactions**](#Nearby_Chromatin_Interactions)
7. [**EndoC/Islet Hi-C**](#EndoC_Hi_C)
8. [**EndoC ChIA-PET**](#EndoC_ChIA_PET)
9. [**EndoC/Islet Chromatin States**](#EndoC_Islet_Chromatin_States)
10. [**EndoC/Islet ATAC-seq**](#EndoC_Islet_ATAC_seq)
11. [**Genes**](#Genes)

**Interactive Tutorial**

In addition to this manual, you can choose to have an interactive tutorial of this app by launching the app and clicking this button near the top left of your webpage:



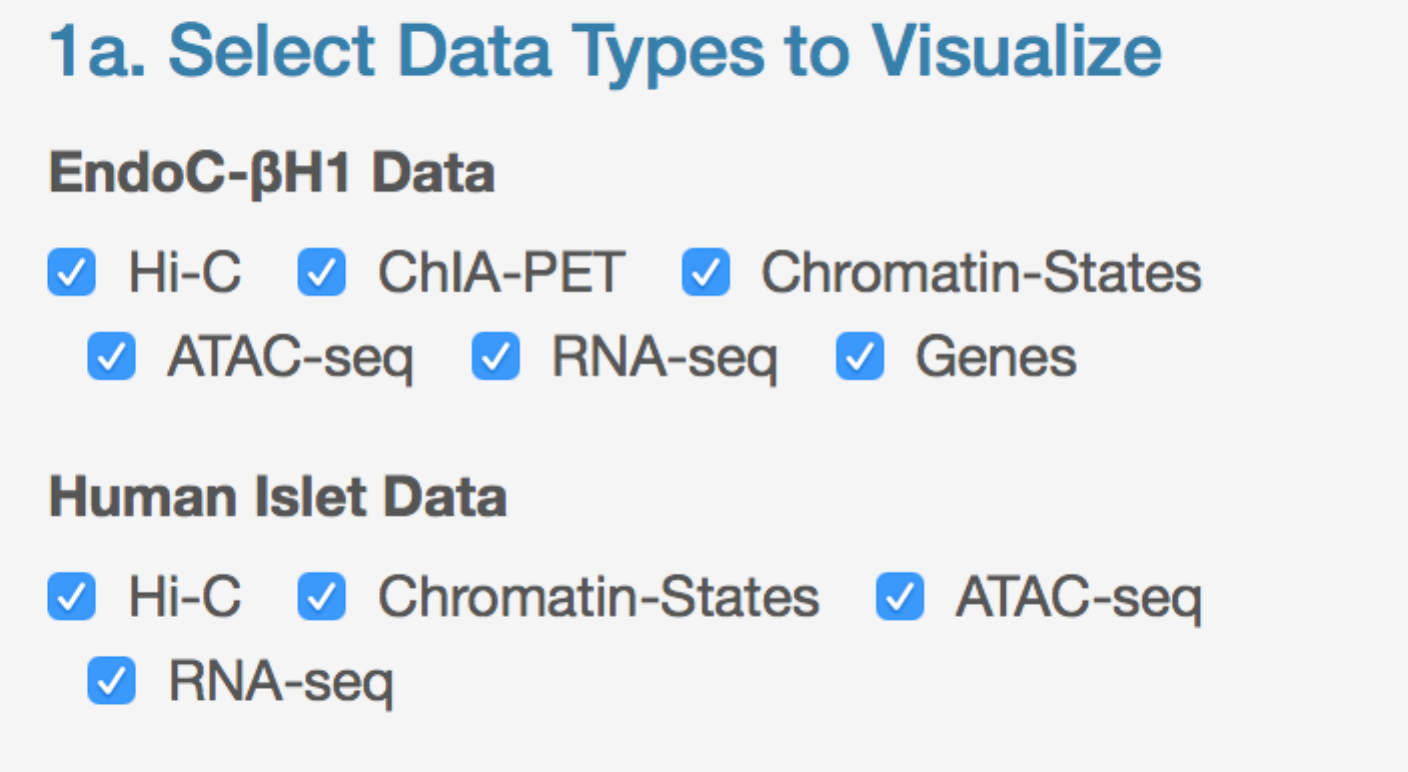
Upon clicking this button, the tutorial should begin and you should see this:

****

Navigate through this tutorial using the “Next Step” and “Previous Step” buttons. To exit out of this tutorial, click the “Quit” button or anywhere outside of the white boxes.

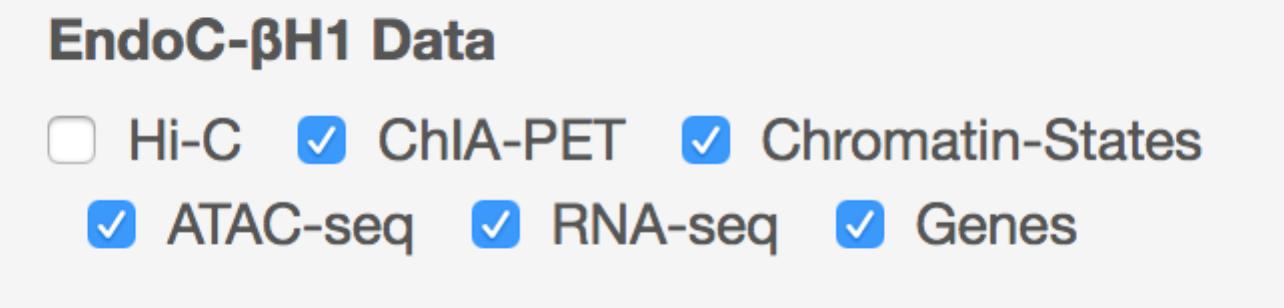
1. **Select Data Types to Visualize**

In the first section of the app, you have the option to specify which data types will be visualized in the plotting window. By default, all data types are selected (for both EndoC and human islet):



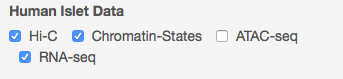
a. EndoC-βH1

For EndoC-βH1, the following data types are available: Hi-C (loop locations), ChIA-PET (Pol2 loop locations), chromatin states (determined via ChromHMM), ATAC-seq (read pileups), RNA-seq (read pileups), and Genes (locations of gencode v19 protein coding and lincRNA genes). To choose to not to display a data type (e.g., Hi-C) in the main plot, simply uncheck the box:



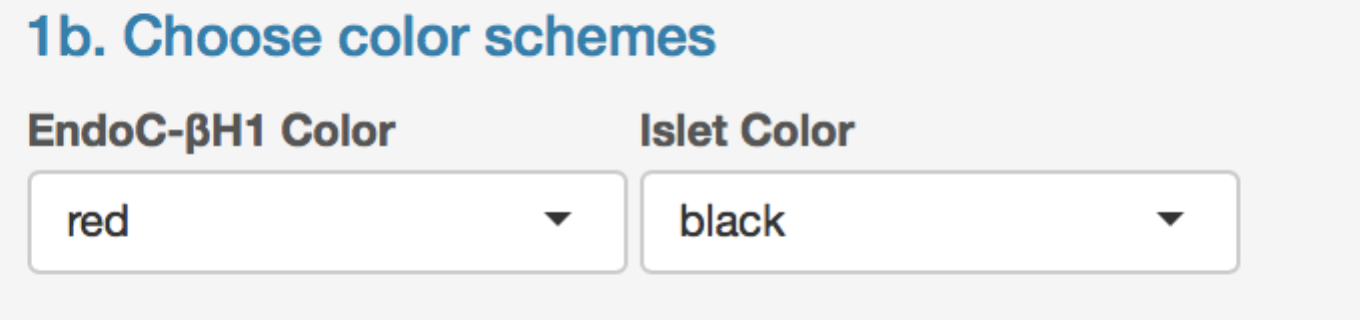
b. Human Islet

Similarly for human islet, multiple data types are available except for ChIA-PET: Hi-C (loop locations), chromatin states (determined via ChromHMM), ATAC-seq (read pileups), RNA-seq (read pileups). Note, there is no “Genes” data type within the Human Islet category because this data is the same for both EndoC and Islets. To choose to not to display a data type (e.g., ATAC-seq) in the main plot, simply uncheck the box:



c. Color schemes

For both data types, the user has the choice of specifying the colors for visualization. Simplify click the drop-down menus and choose your color choice. Default colors are red for EndoC-βH1 and black for Islet.

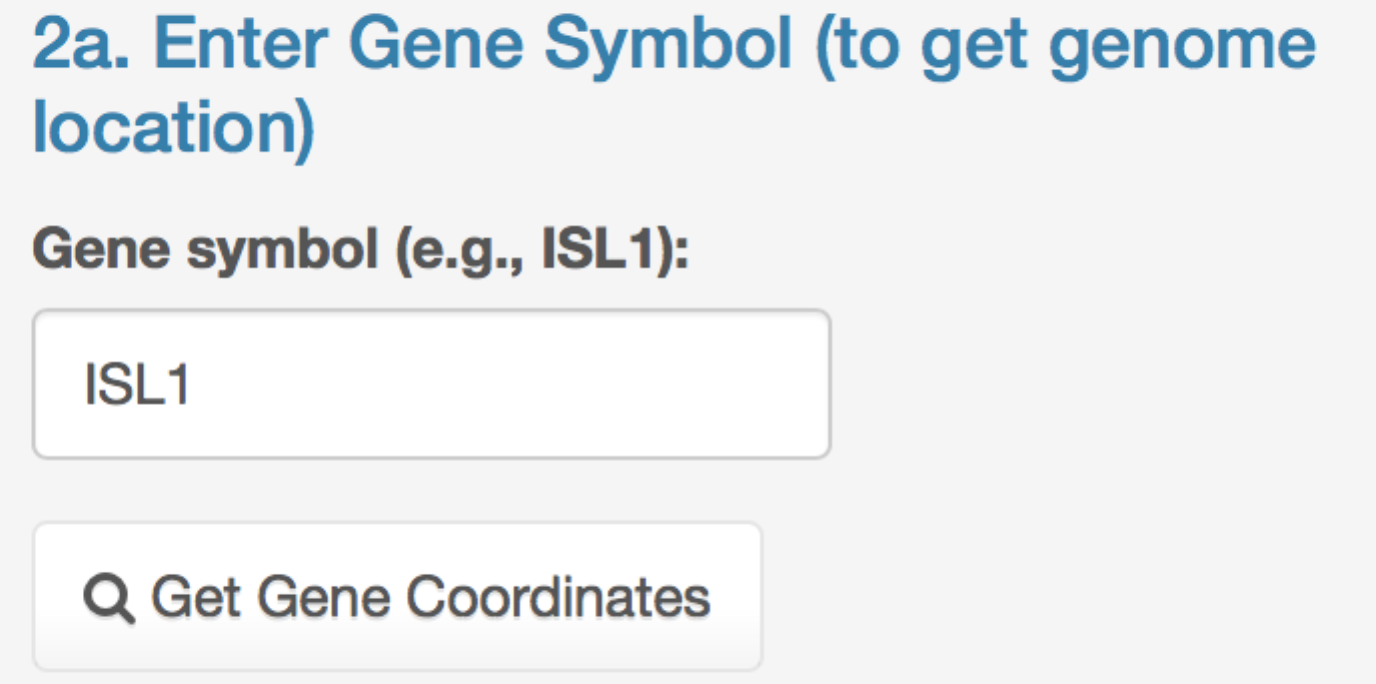


1. **Specifying the Genomic Regions to View**

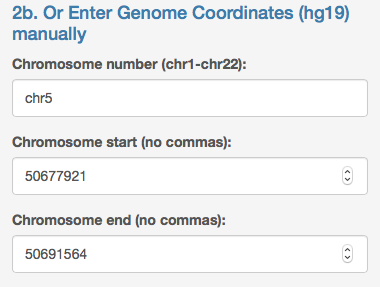
After choosing the data types to visualize, the next step is to choose a genomic region. The app is formatted such that you may focus on a genomic region based on (1) gene positions, or (2) by manually inputted genomic coordinates of interest.

a. Search via gene symbol

To navigate to the genomic position of a particular gene (e.g., ISL1), type in a valid HUGO gene symbol name and click the “Get Gene Coordinates” button.



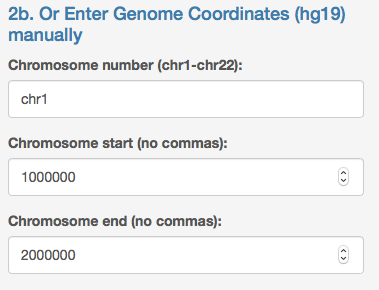
Afterwards, the chromosome number, chromosome start, and chromosome end positions should be updated in the panel below:



Note, if an invalid gene symbol is provided, the chromosome number, start, and end boxes will become “blank”.

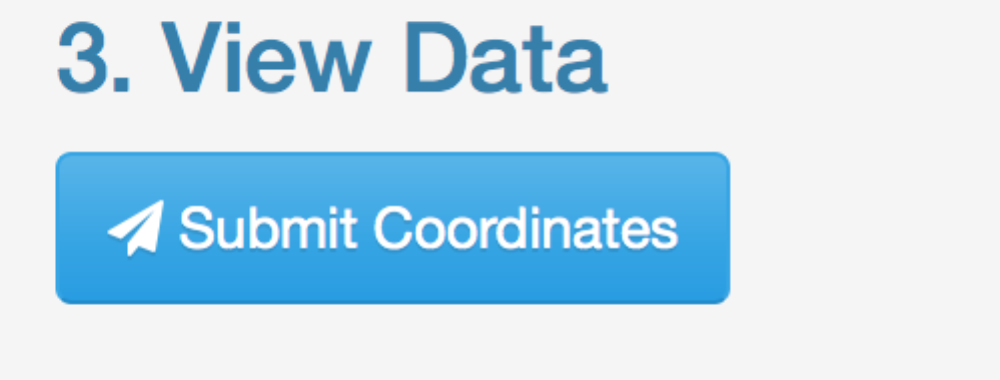
b. Manually specify genomic coordinates

You also may manually enter in a chromosome number, start, and end position. Please note, you may only provide autosomal chromosome numbers (1 through 22) and chromosome start and end positions should be positive integers with no commas. After providing this information, you can now move onto the next section “3. View Data”.

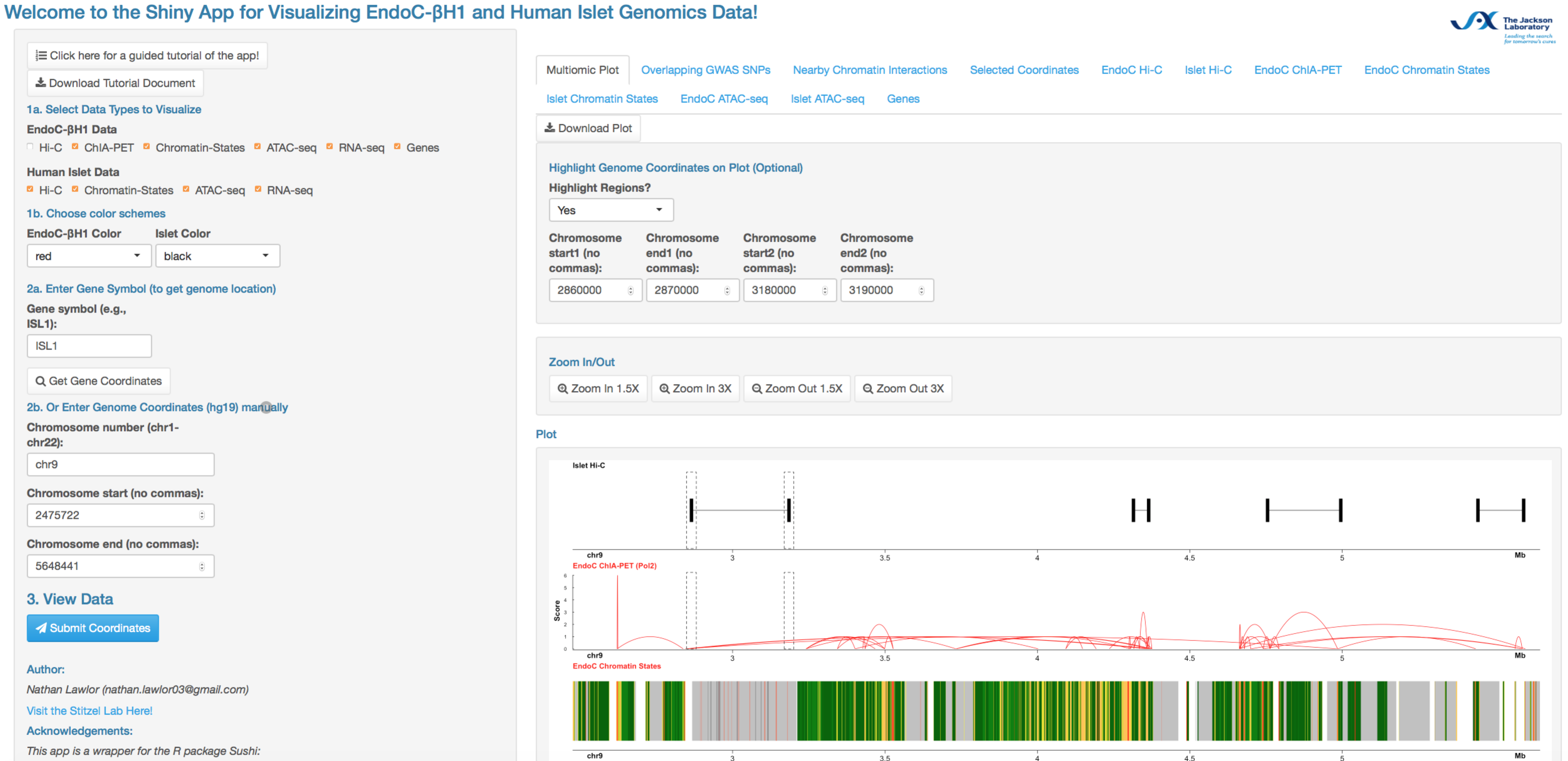


1. **View the Data**

After completing Steps 1 and 2 (required to visualize the data), the next step is to view the data by clicking “Submit Coordinates”. Upon first loading the app, the space underneath the “Plot” area will be blank.



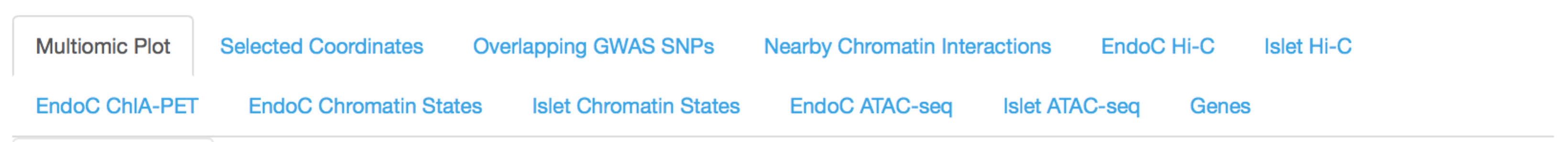
Upon clicking this button, a loading bar should appear in the bottom right of your screen saying “Generating multiomics plot, please wait”. After a few seconds, the plot should be produced in the space to the right of the sidebar (horizontal red arrow) and under the “Multiomic Plot” tab (vertical red arrow) as follows:



After producing this plot, you may use several other functions within the app to further explore and interact with the data. These functions are detailed in the next section Interacting with Tabbed Panels.

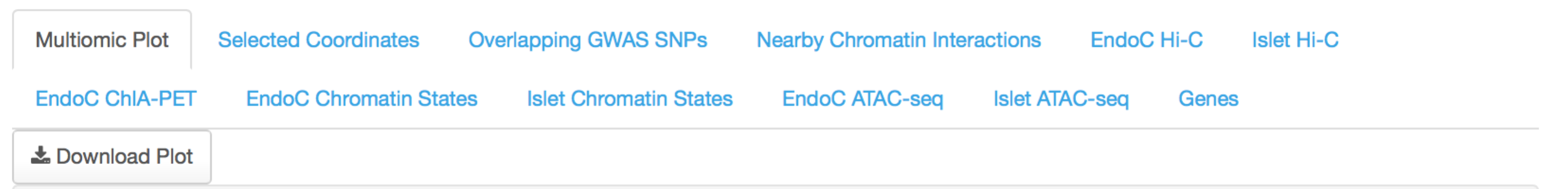
**Interacting with Tabbed Panels:**

After completing Step 3 “View Data” and generating a multiomic plot, many of the tabbed panels located above the main plotting window in the app can be clicked to perform a variety of functions:



1. **Multiomic Plot Panel**

The first tab “Multiomic Plot” can be clicked to view the resulting plot generated in Step 3. Clicking the “Download Plot” button (red arrow) will output a PDF version of the plot displayed in the window.

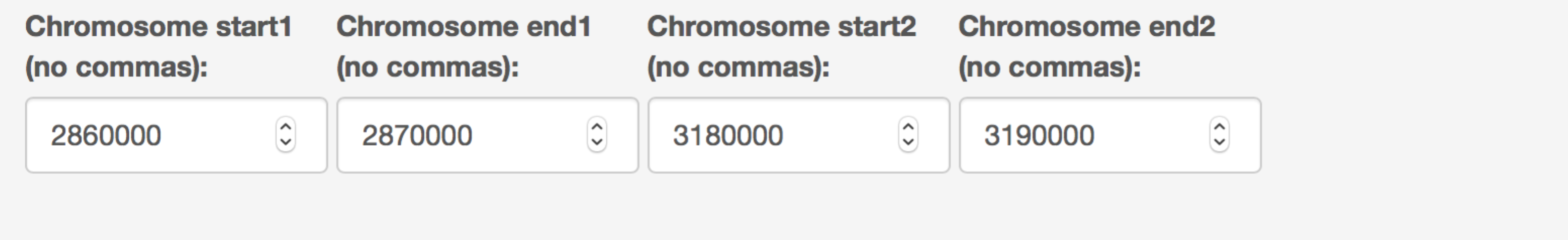


1. **Highlight Regions of Interest**

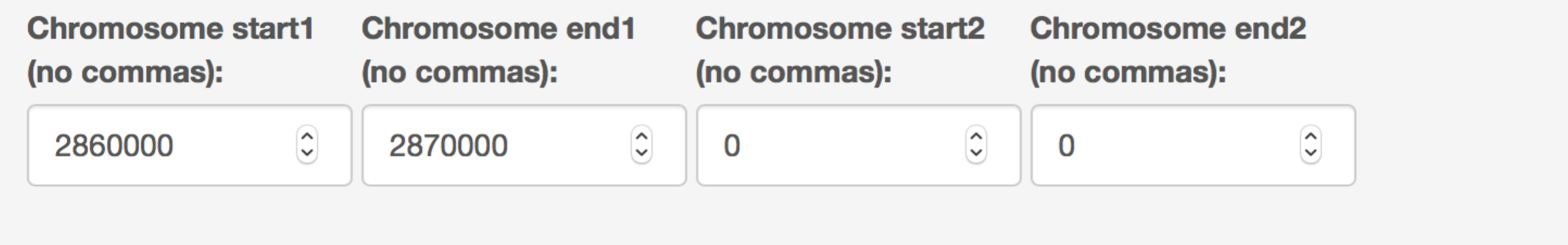
When visualizing genomic regions with this app, you may want to highlight specific genes or features in the overall plot. This step is optional and not needed for the app to run correctly. If desired, you may highlight up to two pairs genomic regions (start position to end position). To do so, first select “Yes” from the “Highlight Regions?” drop-down:



Next, you must manually provide the start and end positions for each pair of regions you want to highlight:

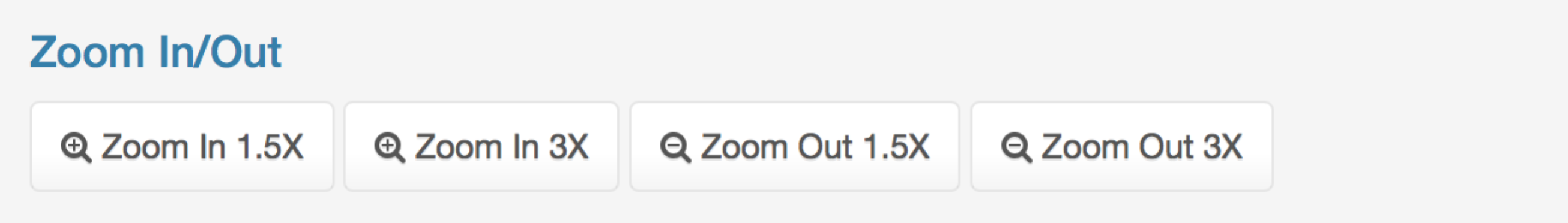


If you would like to highlight only a single genomic region, simply set the positions in “Chromosome start2” and “Chromosome end2” both to zero:

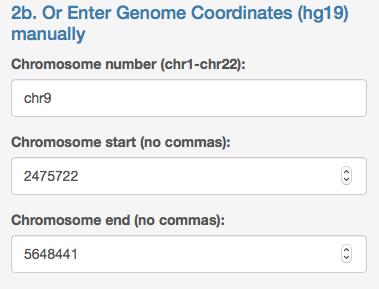


1. **Zoom In and Zoom Out Functions**

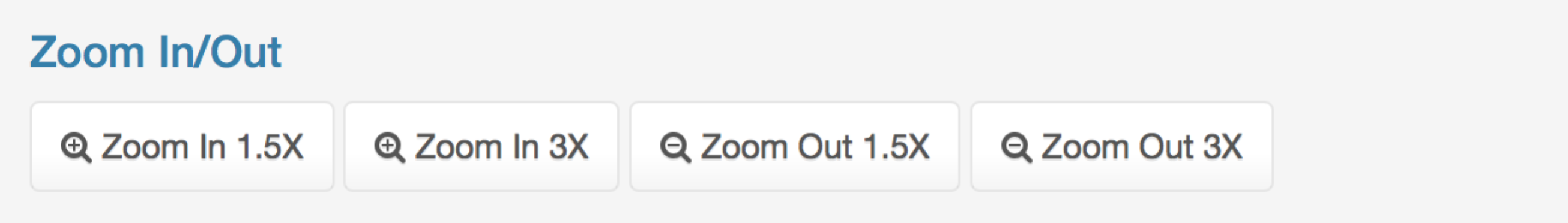
After generating a multiomic plot, you may want to adjust the genomic window you are visualizing. For convenience, “Zoom In” and “Zoom Out” buttons have been provided to zoom in/out at magnifications of 1.5X and 3X:



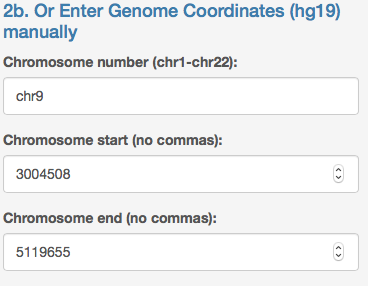
Let’s say you have chosen the following location on chromosome 9 to visualize:



After, viewing the data (Step 5 to be discussed in next section), you realize that you want to zoom in 3X. To do so, simply click the “Zoom in 3X” button once.



Your genome coordinates in Step 2b should now be updated to:



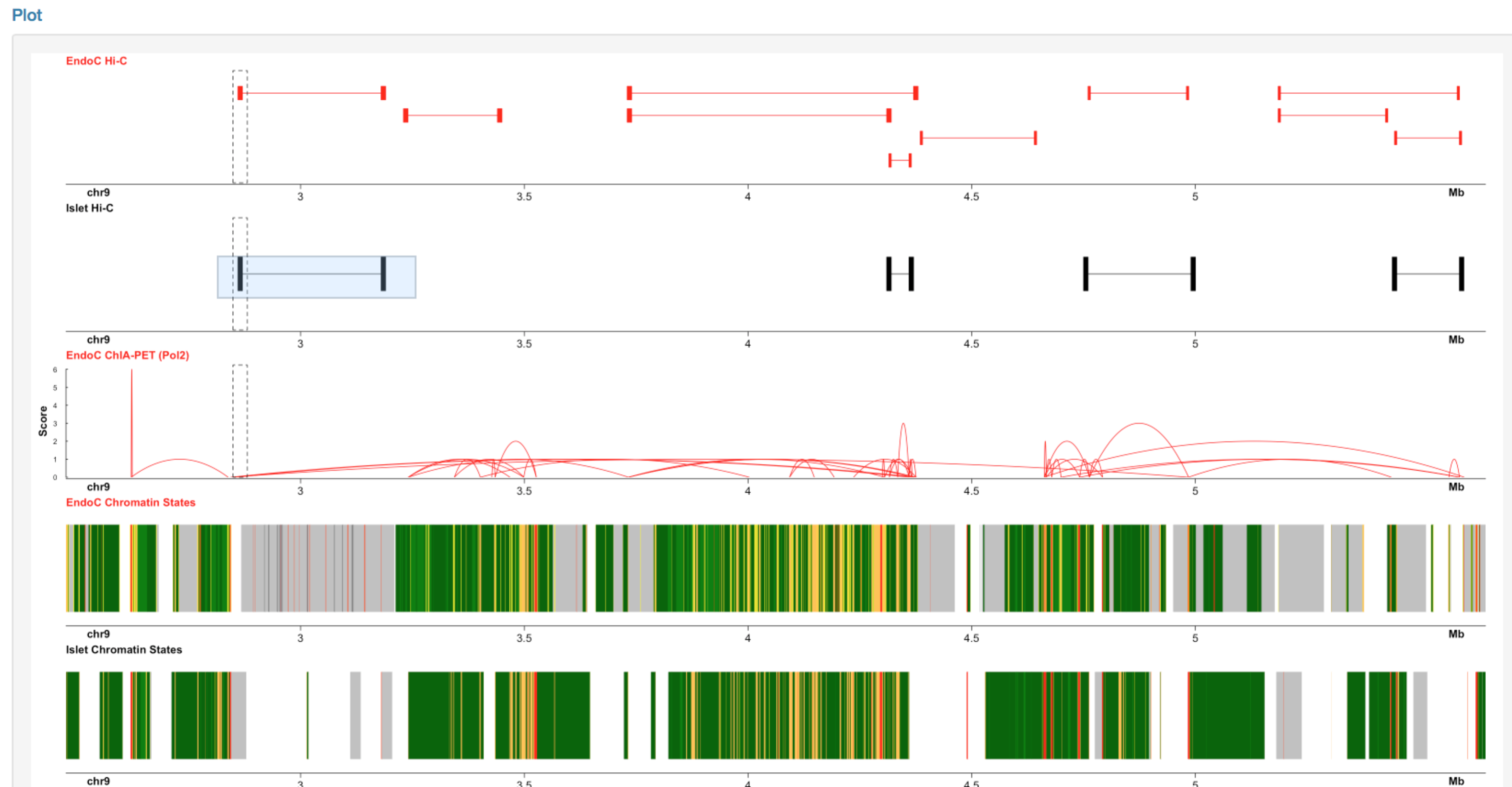
And the plot in the “Multiomic Plot” tab should be adjusted according to these coordinates.

1. **Selected Coordinates**

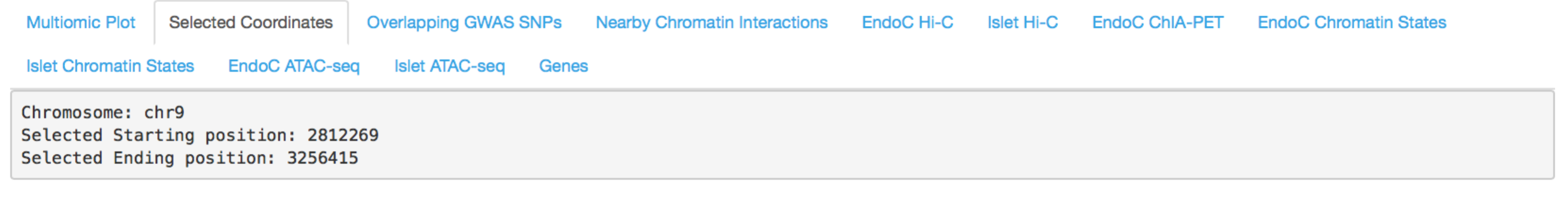
After generating a multiomic plot, this app allows you to click and drag over the plot in the “Multiomic Plot” tab to obtain more information about the features in the region you select. When you first click the “Selected Coordinates” tab, you should see this message:



This indicates that you have not clicked and dragged over the plot in the “Multiomic Plot” tab. To do this, make sure you have clicked the “Multiomic Plot” tab and that you see a plot. Next, move your mouse over the plot image until you see the mouse cursor change into a “+” symbol. Then, click and drag your mouse over the region you wish to select:



Afterwards you should see a blue box (emphasized by red arrow) indicating the region you selected. If you then click the “Selected Coordinates” tab again, you should now see:



Which indicates the chromosome number, starting position, and ending position of the region you just selected. Note the height of blue box is not important (it doesn’t matter if you selected box spans the entire height of the plot or only a single data track. Rather, only the width of the blue box is used to identify a selected starting and ending position.

After selecting a region on the plot and clicking any of the following “Overlapping GWAS SNPs” through “Genes” tabs:



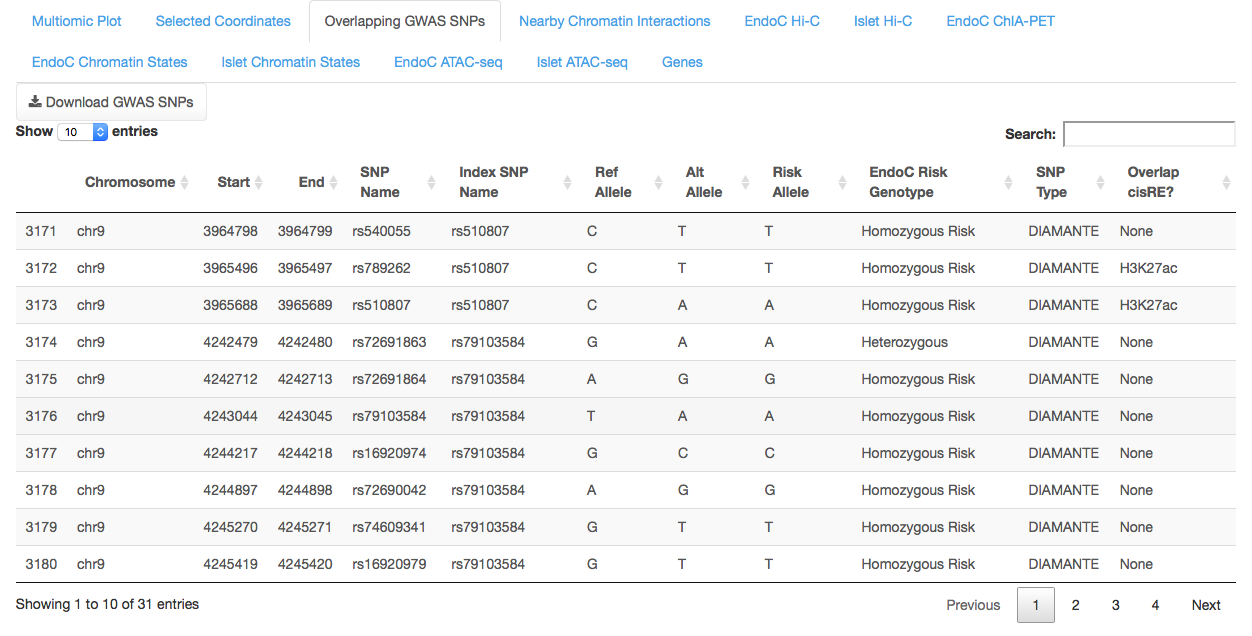
The resultant tables in each tab will show you the features present in the selected region.

To remove a selection on your plot, simply double click anywhere on the plot and ensure the “blue box” disappears. To verify the selection has been removed, you can check the “Selected Coordinates” tab again which should no longer show chromosome number, start, and end information.

When no regions are selected on the plot, the tabs “Overlapping GWAS SNPS” through “Genes” tabs will show all features present in the genomic region specified by the coordinates in Panel 2b.

1. **Overlapping GWAS SNPs**

Upon clicking the second tab “Overlapping GWAS SNPs”, you can view Type 2 diabetes and other metabolic-associated genome wide association study (GWAS) single nucleotide polymorphisms (SNPs) that intersect the genomic region you are viewing:



The interactive table above shows GWAS SNPs that overlap chr9: 2475722- 5648441. Each row of the table represents a single SNP and contains the SNP location, rs identifier, lead index SNP name, hg19 reference allele and alternate allele at the SNP position, the risk allele information in EndoC cells, EndoC genotype, disease category, and whether or not the SNP overlaps an EndoC cis-regulatory element (cisRE). Clicking any of the columns in this table will sort the results in ascending/descending order. Similarly, you can search by features (e.g. SNP\_Name, etc.) in the Search box to better filter your results.

Note, selecting a region on the multiomic plot as previously described, will filter results in this table to those within the selected region (rather than the entire plot window).

Click the “Download GWAS SNPs” button near the top left of the table (red arrow) to obtain a CSV file of the current table in view (containing all results in the entire plot window or the results within the selected region you specified by clicking/dragging over the plot). The output file name will specify the genomic coordinates considered in the table.

1. **Nearby Chromatin Interactions**

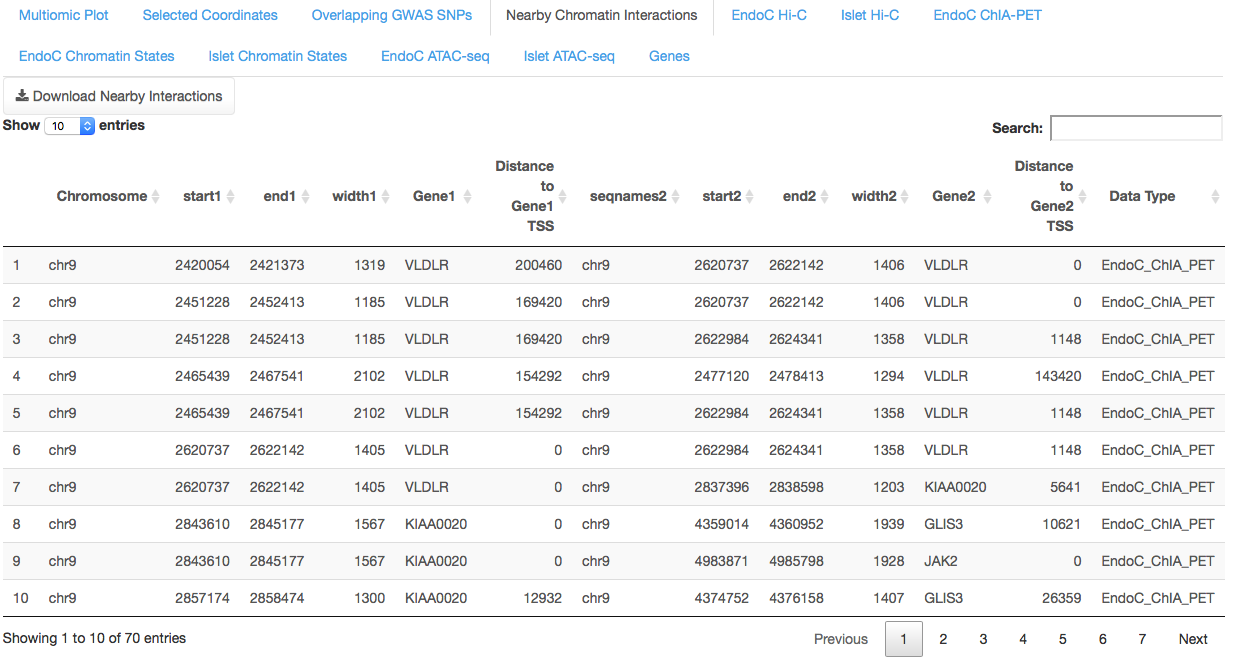
Chromatin interaction data (Hi-C, ChIA-PET) are often represented as a pair of genomic coordinates (also called anchors). For example, in the following ChIA-PET interaction:



The first anchor is at chr9 spanning from 2420054 – 2421373 and the second anchor is at chr9 spanning from 2620737 – 2622142.

In order to visualize chromatin interactions within the app, both anchors must be located within the genomic coordinates specified in Step 2b. Say the genomic coordinates we specify are: chr9:2475722 – 5648441. Because the first anchor of this ChIA-PET interaction occurs outside of the region we specified (even though the second anchor occurs within the region we specified), this interaction will not be displayed on our multiomic plot.

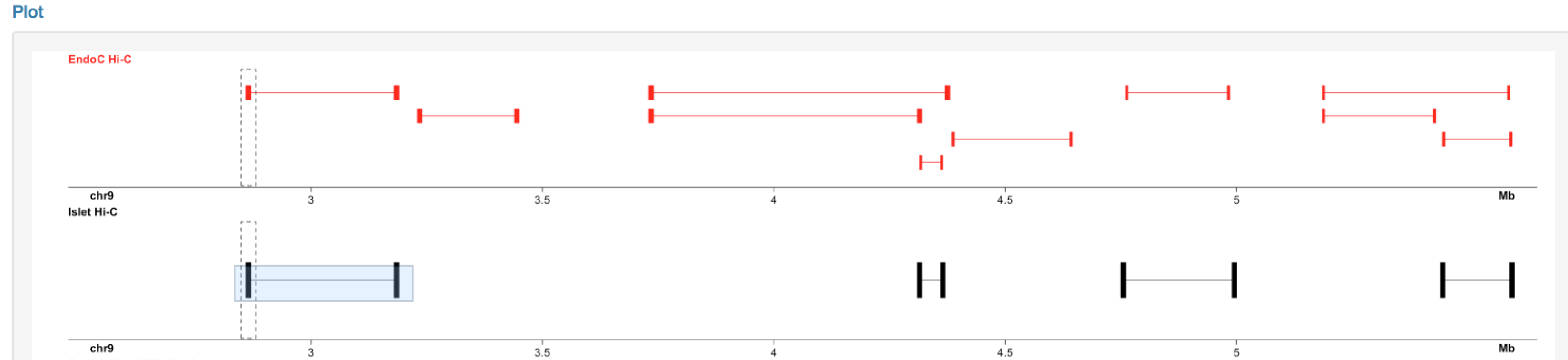
By clicking this “Nearby Chromatin Interactions” tab, you are given a table:

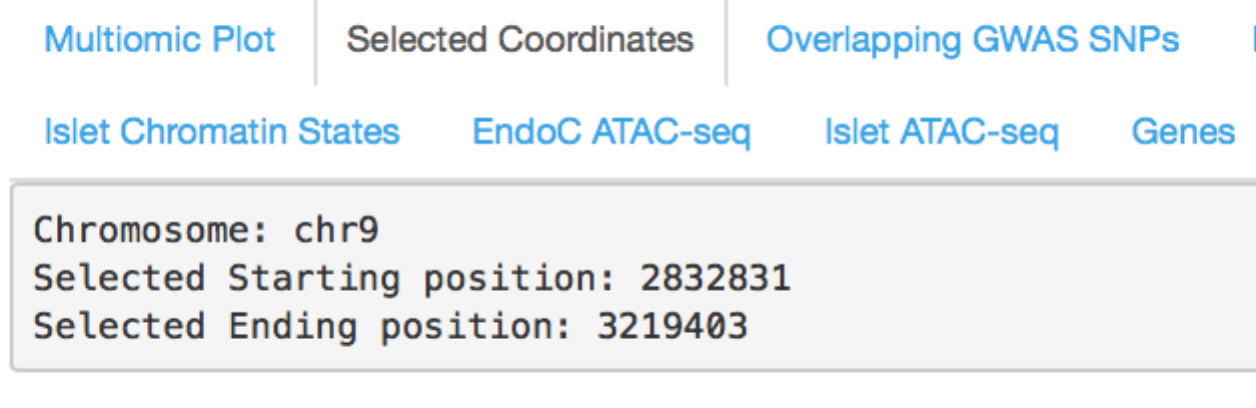


Which contains all chromatin interactions (ChIA-PET, Hi-C) that contain at least one anchor that intersect with the genomic coordinates specified in Step 2b (including the example above that was not plotted). This table is helpful for identifying chromatin interactions that may be spanning upstream or downstream of your plotting window. For each anchor, the nearest gene annotation and distance to the gene transcription start site (TSS) is provided.

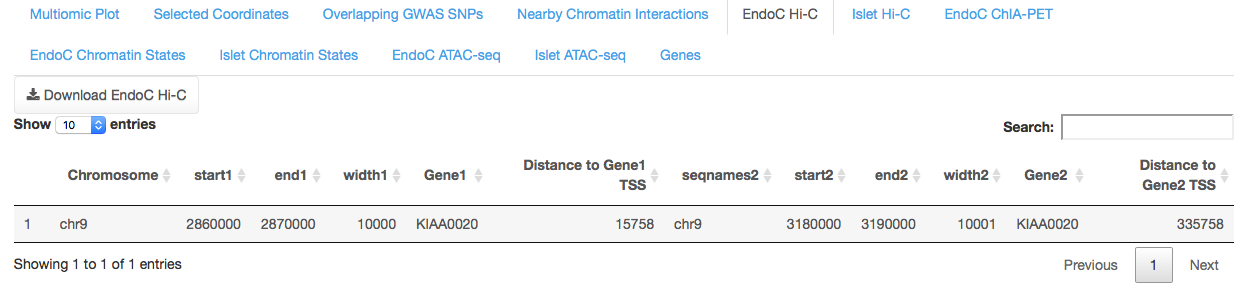
1. **EndoC/Islet Hi-C**

After selecting coordinates in Step 9, let’s say the region you selected is at a region on chr9:



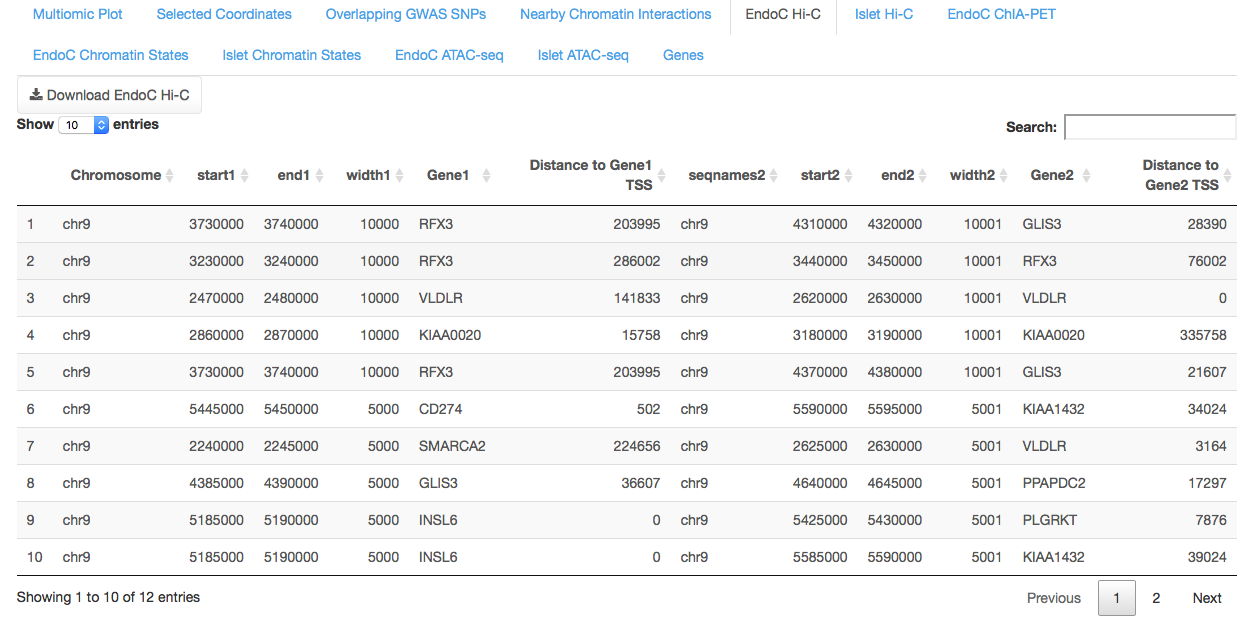


Next, if you click the “EndoC Hi-C” (or the “Islet Hi-C”) tab, you will see a table in the following format:



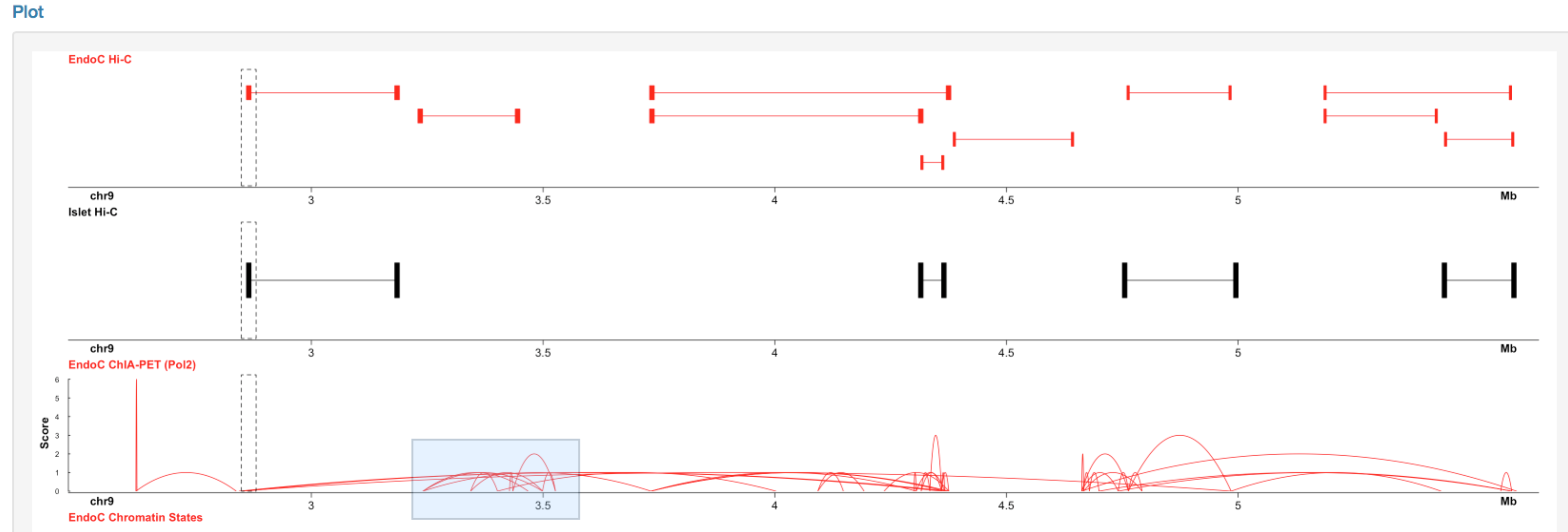
Which indicates any Hi-C chromatin interactions in EndoC (or Islet) cells that occur within the region you previously selected. If no chromatin interactions are present, the table will be empty. Clicking the “Download EndoC Hi-C” button will download a CSV file of this table.

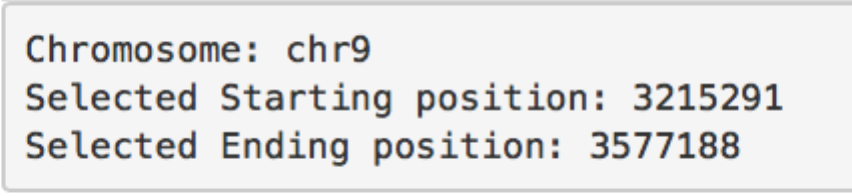
If you were to not select any regions in the above plot, the “EndoC Hi-C” (or “Islet Hi-C”) tab should display all EndoC (or Islet) Hi-C loops within the plotting window:



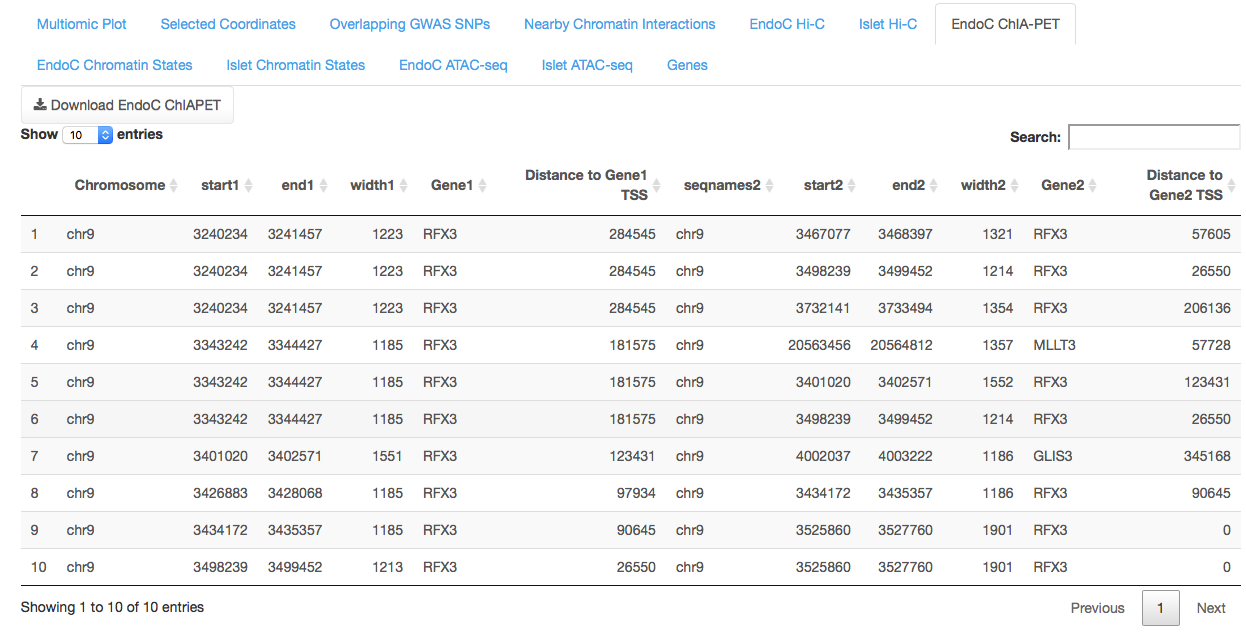
1. **EndoC ChIA-PET**

Now, let’s say you select another region on chr9:





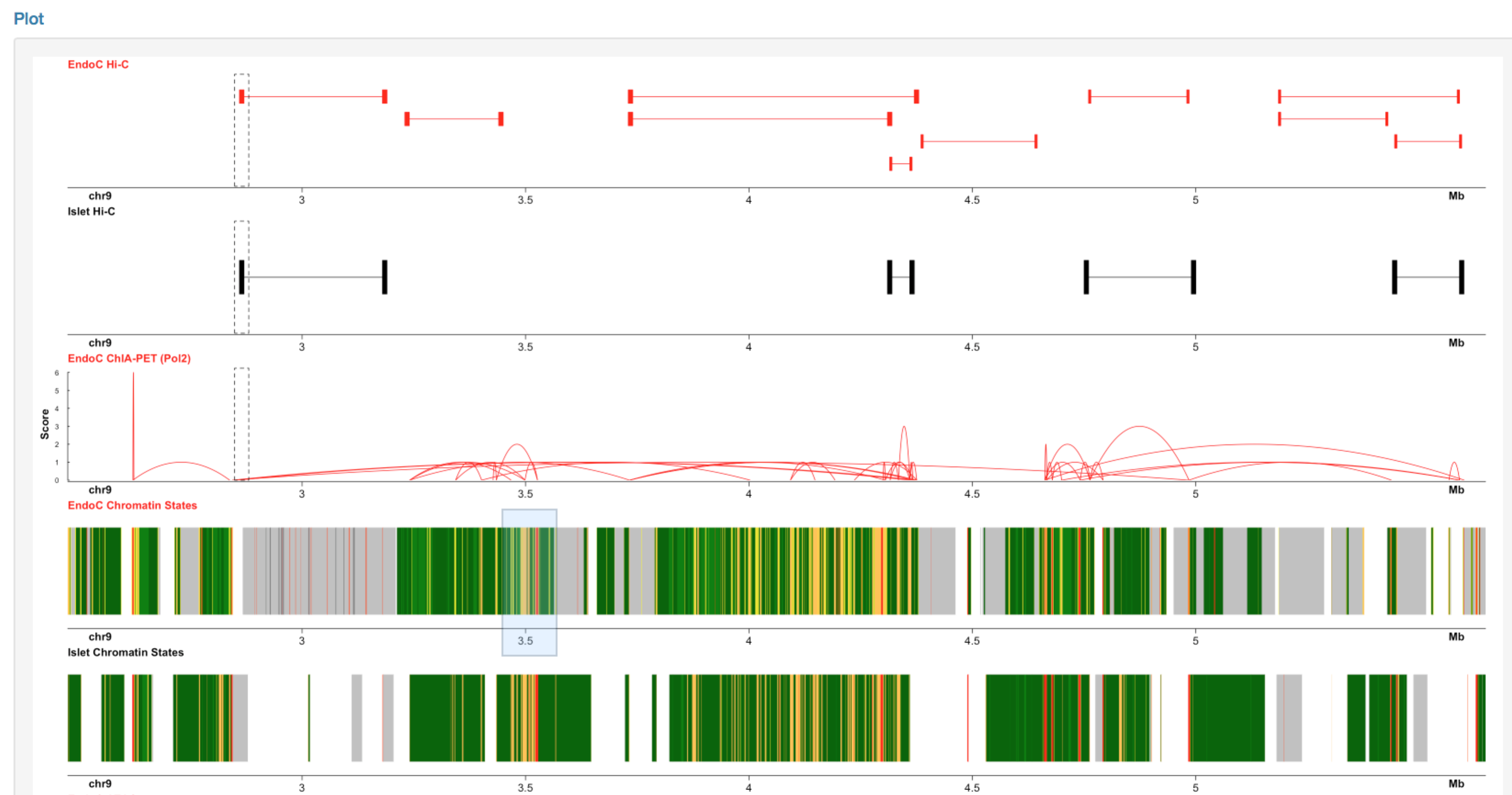
Next, if you click the “EndoC ChIA-PET” tab, you will see a table:



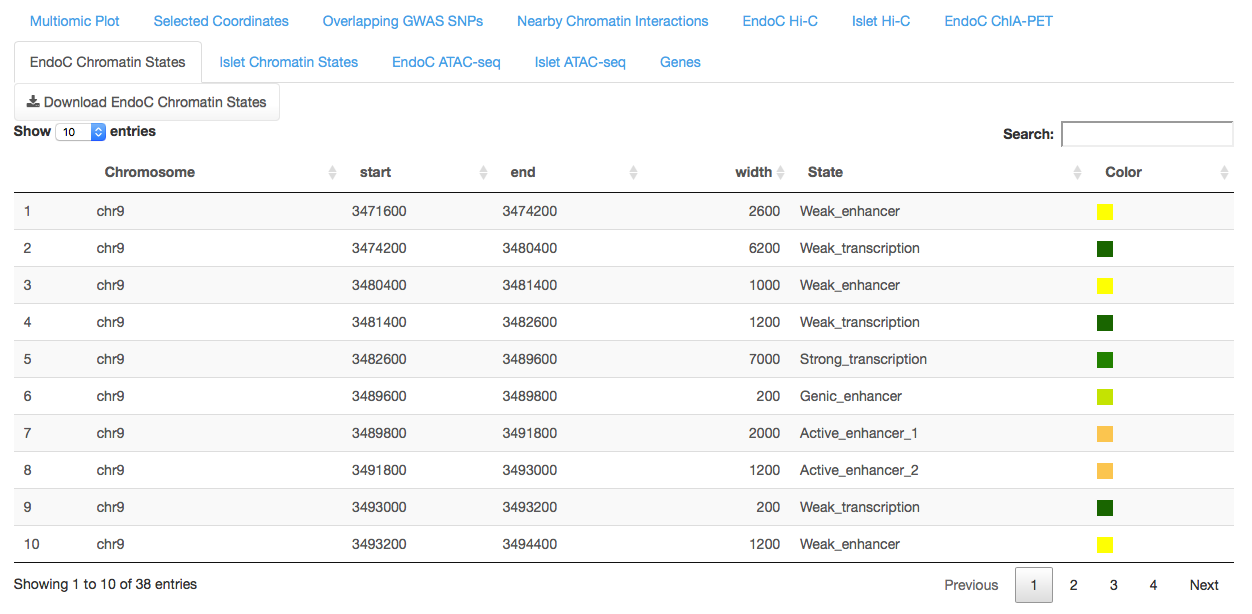
Which indicates any ChIA-PET (Pol2) chromatin interactions in EndoC cells that occur within the region you previously selected. If no chromatin interactions are present, the table will be empty. Clicking the “Download EndoC ChIAPET” button will download a CSV file of this table.

1. **EndoC/Islet Chromatin States**

For a selected region on chr9 (emphasized by red arrow):



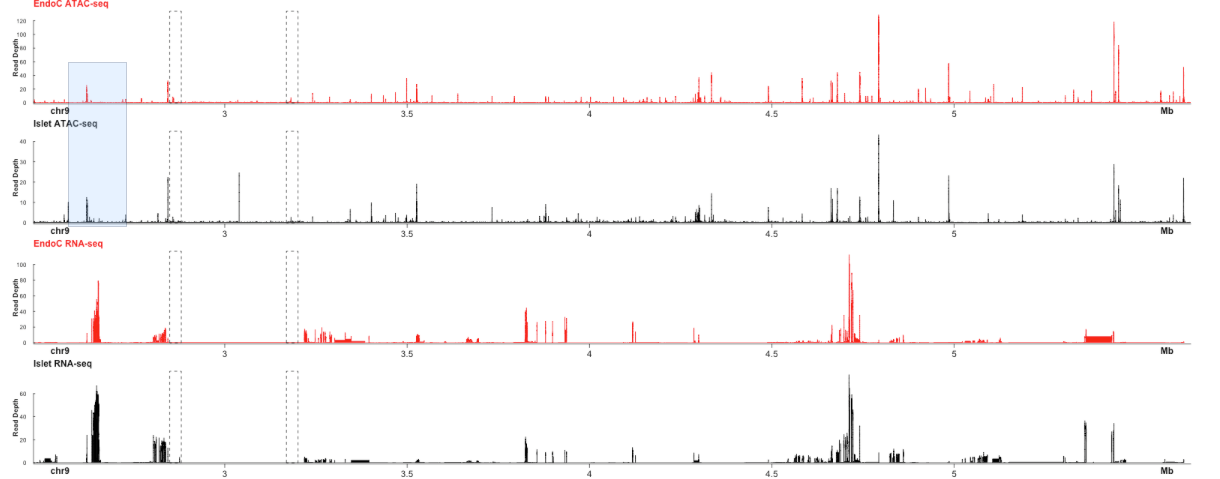
Click either the “EndoC Chromatin States” or “Islet Chromatin States” tabs. You should see the following type of table:



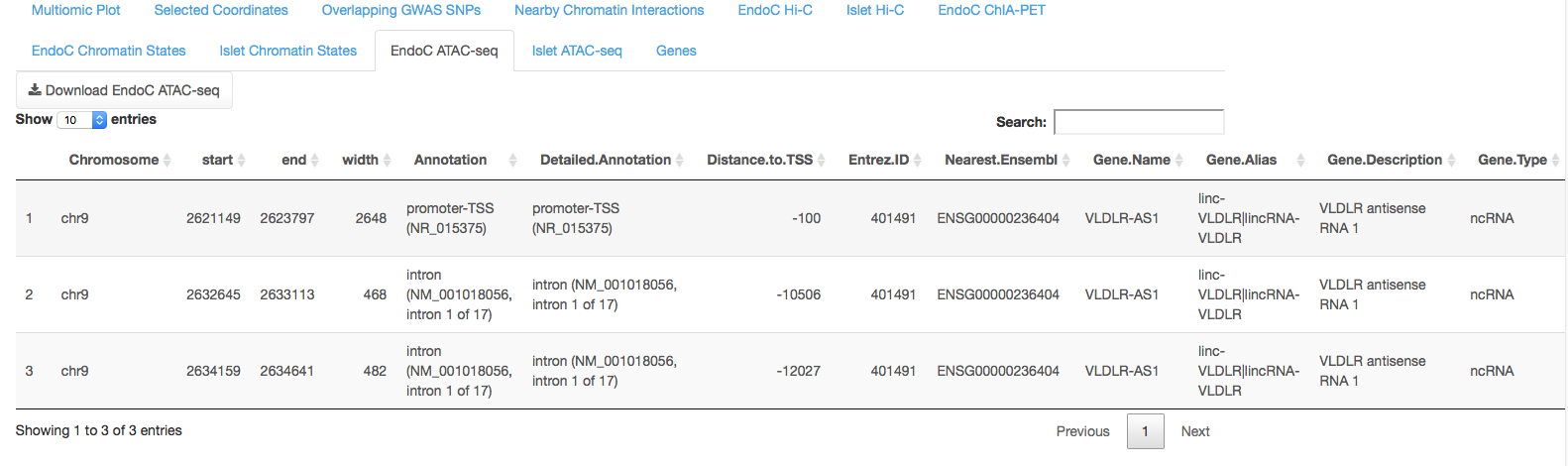
Which will indicate the positions of each chromatin state, the state annotation, and the corresponding color. If no chromatin state information is available for the selected coordinates, the table will be empty. A Download button should be located in the top left corner for each table that will output a CSV file of the current table.

1. **EndoC/Islet ATAC-seq**

For a selected region on chr9:



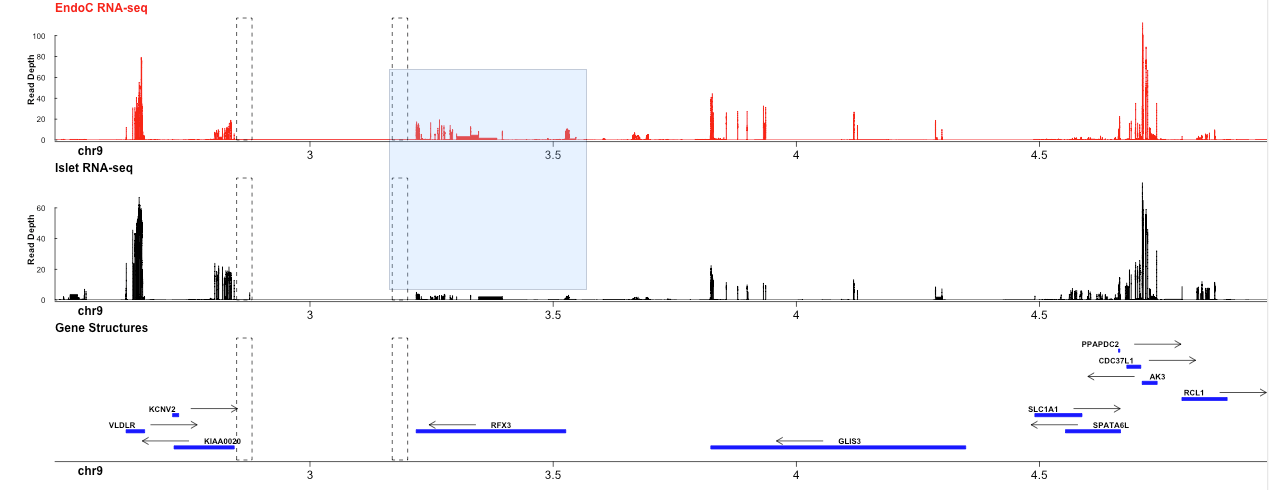
Click either the “EndoC ATAC-seq” or “Islet ATAC-seq” tabs. You should see the following type of table:



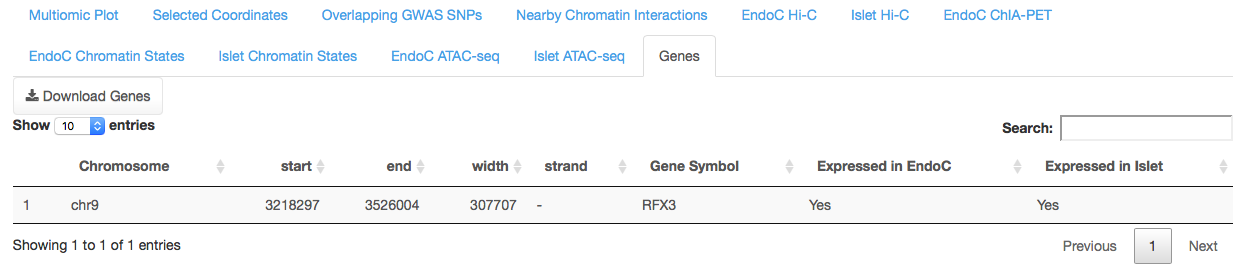
Which will indicate the positions of called ATAC-seq peaks (using the MACS2 algorithm), and the HOMER annotations for the peaks (nearest gene name, distance to nearest gene TSS, gene description, gene type, etc.). A Download button should be located in the top left corner for each table that will output a CSV file of the current table.

1. **Genes**

For a particular region on chr9:



Next, click on the “Genes” tab, which should display a table like:



Which contains the positions of any genes that occur within the selected region, and whether or not the gene is expressed in EndoC or Islet RNA-seq data. A gene is considered expressed with a fragment per kilobase of transcript per million mapped reads (FPKM) value ≥ 1. Click the Download Genes button in the top left corner to obtain a CSV file of the current table.