WashU EpiGenome Browser Tutorial

March 28, 2019 – City of Hope.

Presenter: Ting Wang

Tutorial Overview:

Notes:

- To follow along with this tutorial, use instructions marked by ➢. We have provided screenshots for guidance (ordered by bracketed alphabets: [A], [B], …)
- Click in the order of the numbered-circles on the screenshots to follow this tutorial.

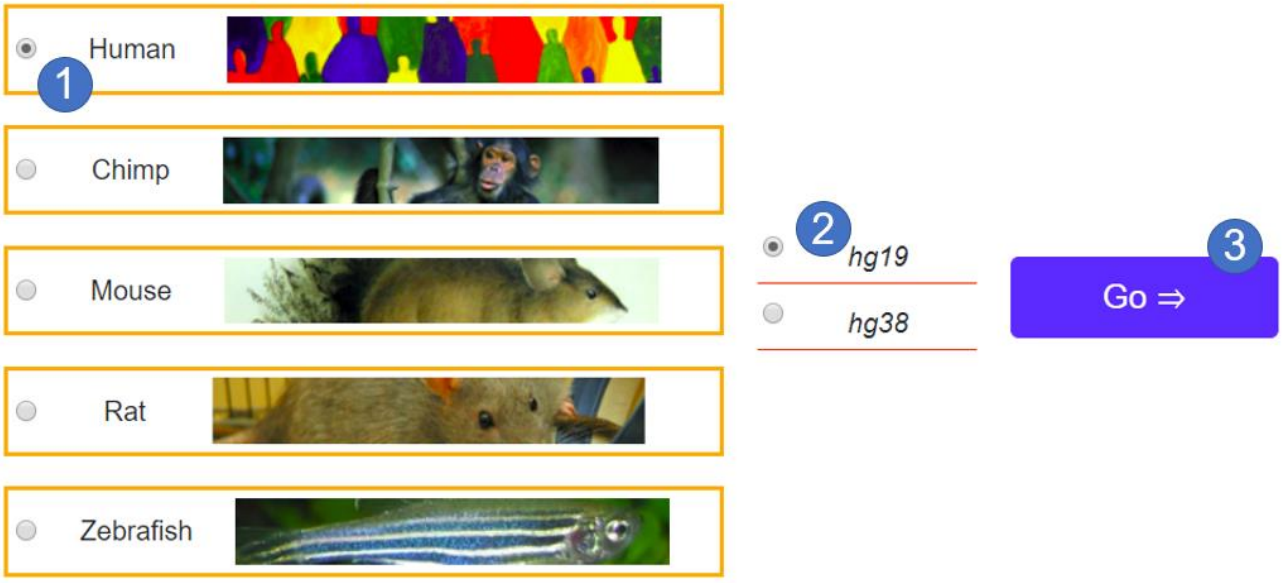
1. Getting started with the EpiGenome Browser

1.1. Load the browser

- Access the browser at http://epigenomegateway.wustl.edu/browser

1.2. Select the genome assembly of interest

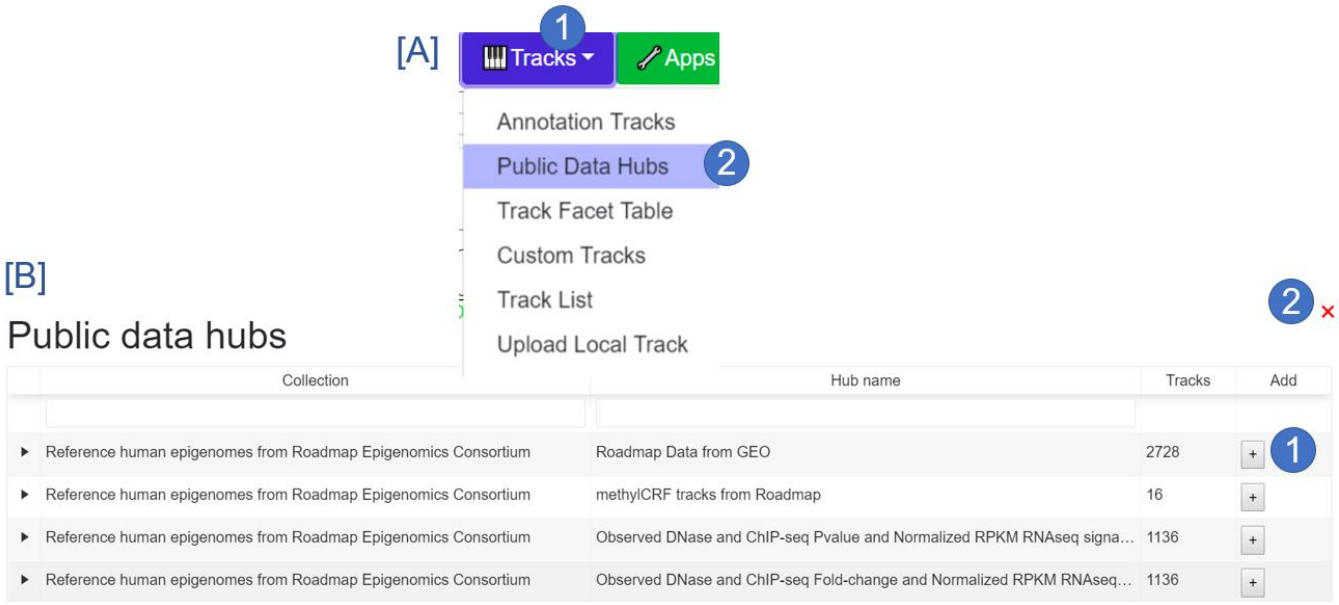
- For the purpose of this tutorial, please select **Human hg19**.



2. Loading data on the EpiGenome Browser

2.1. **Data hubs:** A data hub is a collection of tracks/datasets that can be viewed on the browser.

- [A] Click on the **Tracks -> Public Data Hubs (21 available)** to view all the available public datasets on the EpiGenome Browser.
- [B] Choose the first hub named **Roadmap Data from GEO**, Click the **Add** button. **Exit** the data hub section by clicking on the **X** at the top-right of the floating window.

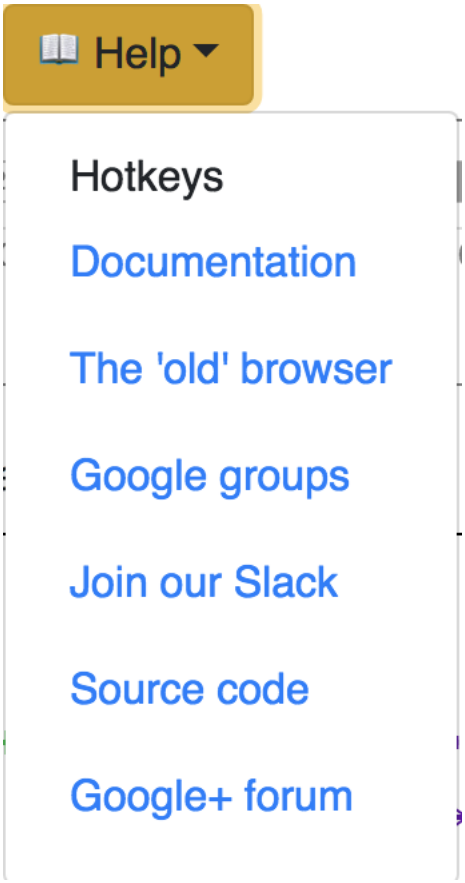


3. Navigating the EpiGenome Browser

3.1. Layout of the EpiGenome Browser

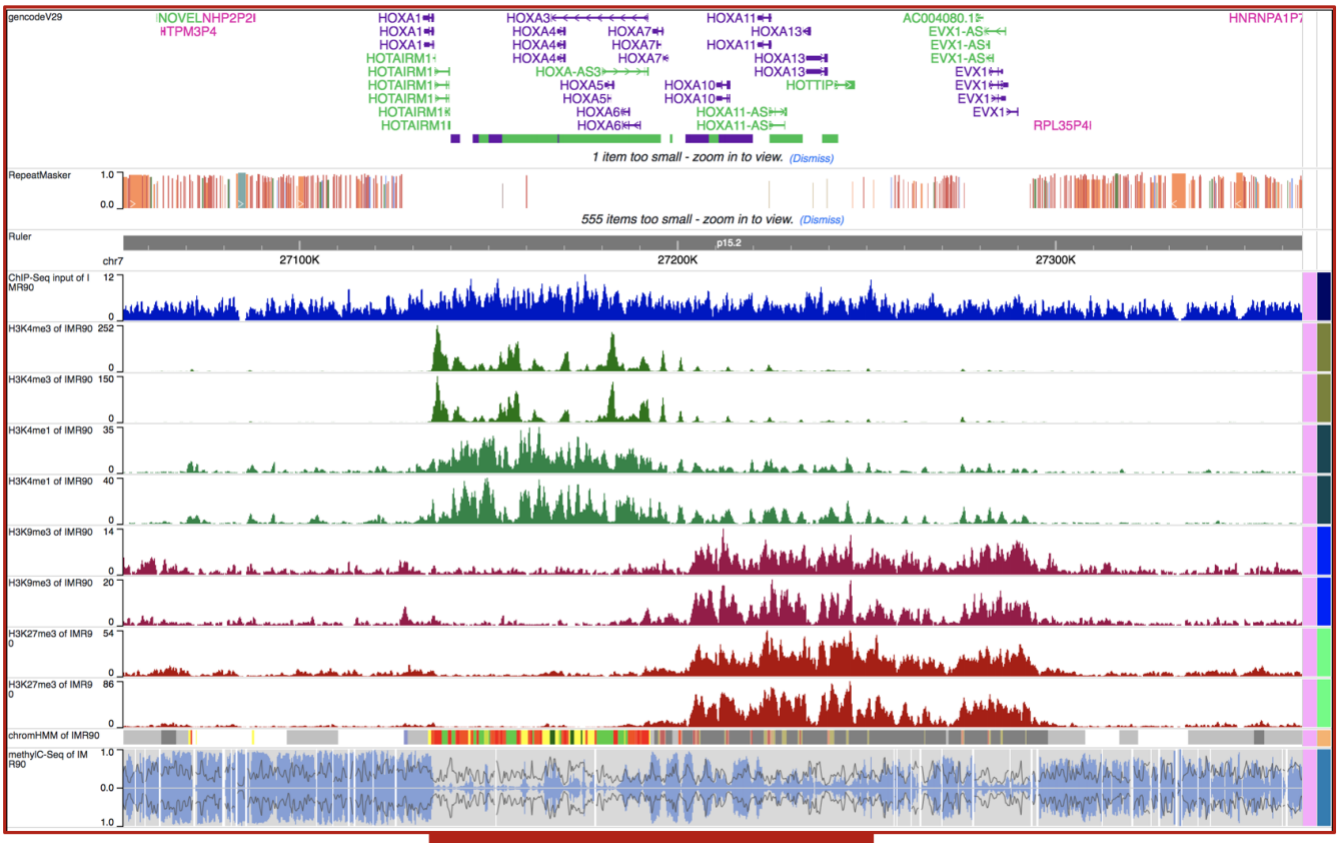
3.1.1. Help

- This menu contains links to: (i) hotkeys information, (ii) the documentation on the browser, (iii) link to the 'old' browser and (iv) discussion forums (on Google groups, Slack channel, Google+ and Facebook, as well as source code repository) for the browser.



3.1.2. EpiGenome Browser

- This region contains the data tracks aligned to the genome, overlaid on gene annotations and other annotations customized by the user.

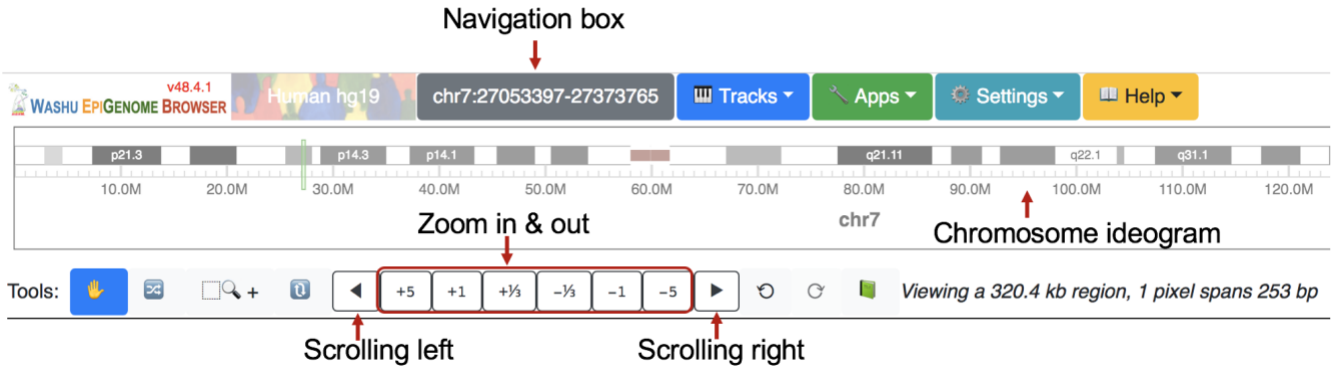


WashU Epigenome Browser

3.2. Navigating the genome on the EpiGenome Browser

3.2.1. Genome navigation controls

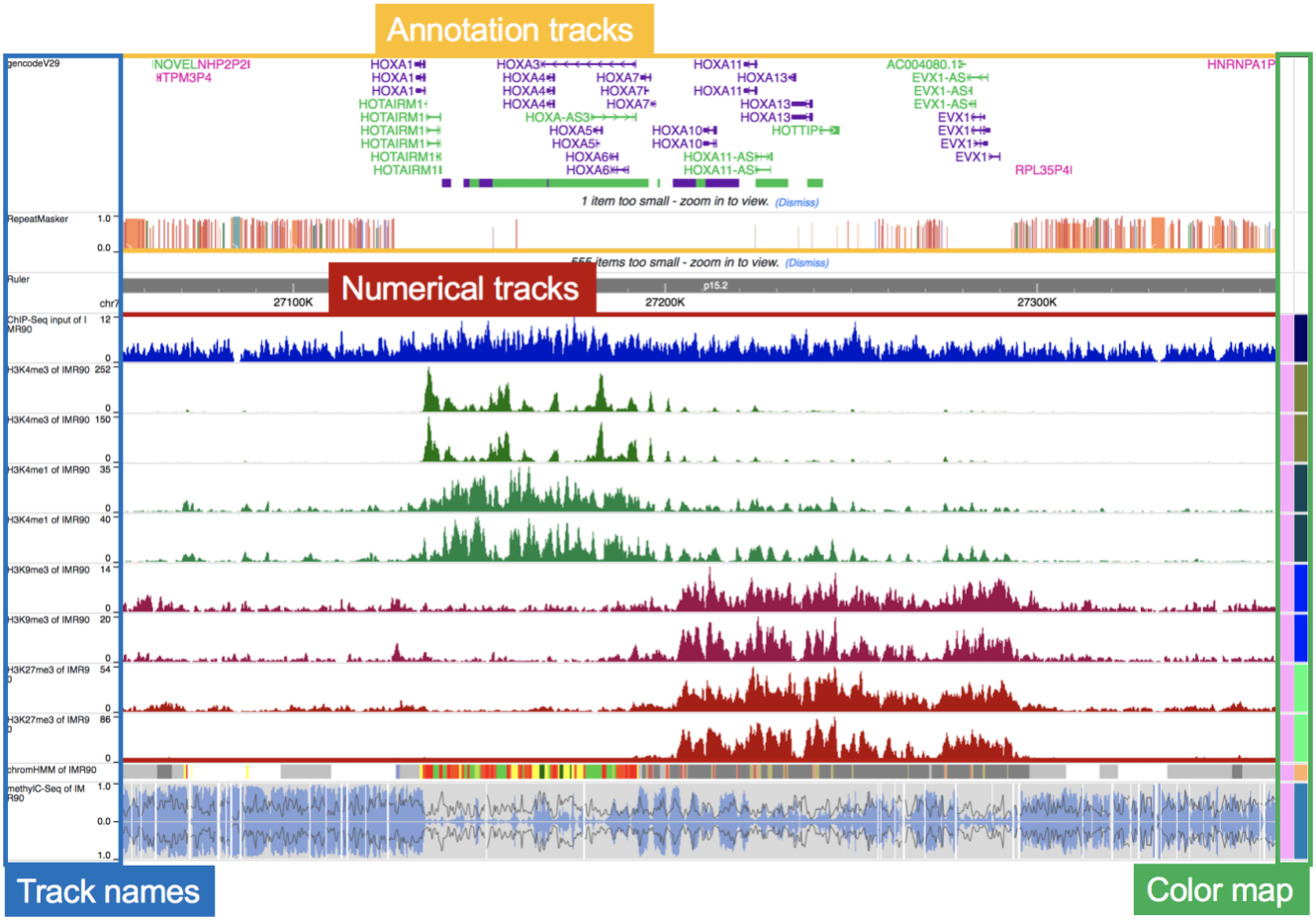
- The controls enable moving across the genome, using the **zoom-in**, **zoom-out**, and **scrolling** buttons.
- Alternatively, click on the **navigation box** to enter the genomic coordinates of the region of interest. Click on **Go** to move to the new region.



3.2.2. Data visualization

- **Numerical tracks** represent the density of reads from sequencing experiments aligned to the genome assembly.
- **Track names** for each track are listed on the left-hand side of the track itself.

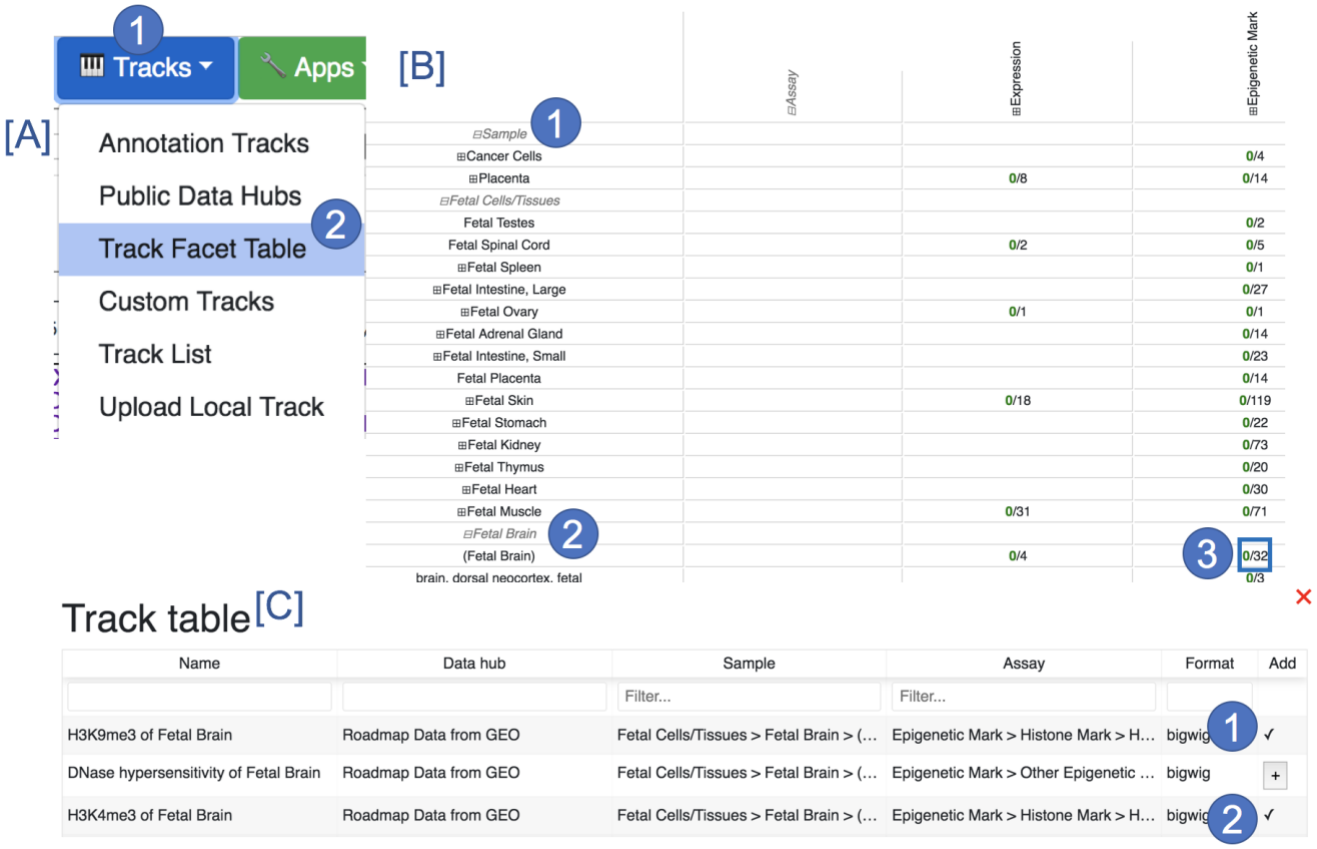
- On the right-hand side of the tracks is a **color map**, which represents the **metadata** for the tracks.
- Above the numerical tracks are **annotation tracks**, including annotations of genes and transposable elements.



3.3. Searching for datasets on the EpiGenome Browser

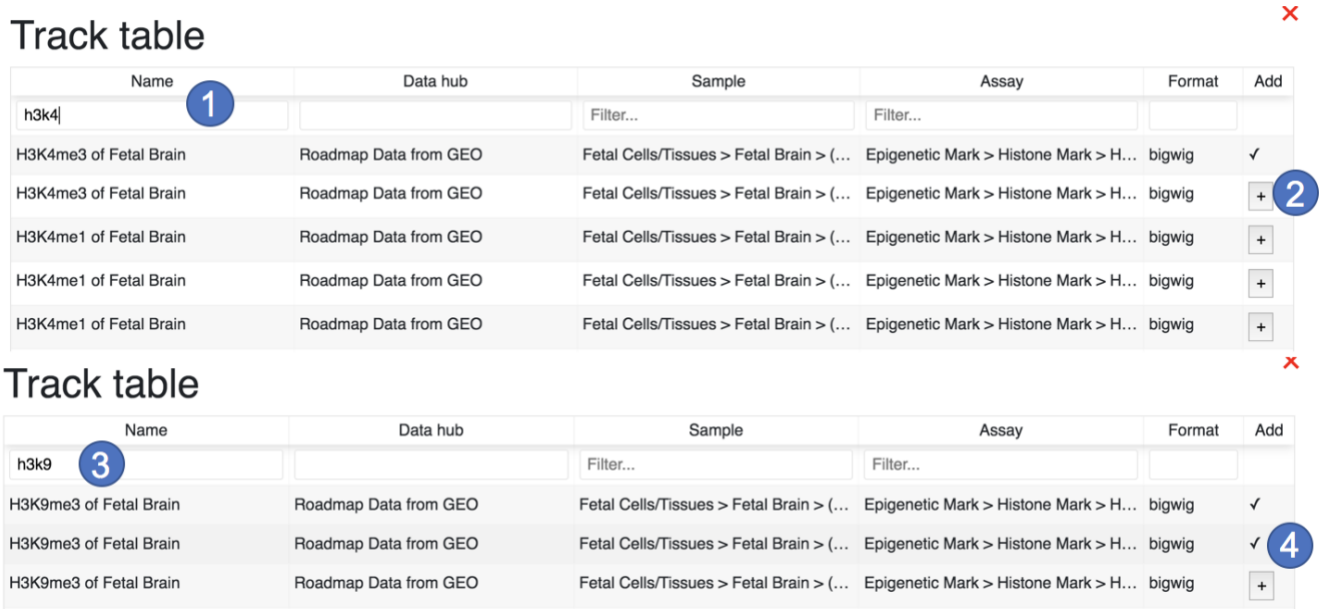
3.3.1. **Facet table:** The facet table organizes all the loaded datasets into a table, to allow the user to search for datasets using metadata terms. Here, **rows** represent samples and **columns** represent assay types. **Cells** with numbers represent the number of datasets. For example, the cell with numbers: **11/602** - here, **11** represents the number of datasets currently loaded on the browser, while **602** represents the total number of datasets available.

- [A] To display the facet table click on the **Tracks** menu, and then click on the **Track Facet Table** box.
- [B] This will generate the facet table. To add more datasets, click on **Fetal Cells/Tissues** and then select **Fetal Brain**. Click on the cell corresponding to the **Epigenetic Mark** column and the **Fetal Brain** row to list the datasets in this category.
- [C] To select data, click the **Add** button of **H3K9me3 of Fetal Brain** and **H3K4me3 of Fetal Brain**.
- To **exit** this floating window, click on the **X** at the top-right of the window.



3.3.2. Searching for track in track table

- Click on the cell at Fetal Brain row and Epigenetic Mark column on the facet table.
- This will list all the dataset under this 2 metadata terms. Type h3k4 in the box under Name header. Add the track **H3K4me3 of Fetal Brain**.
- Repeat this process to add one more replicate track for **H3K9me3 of Fetal Brain.**

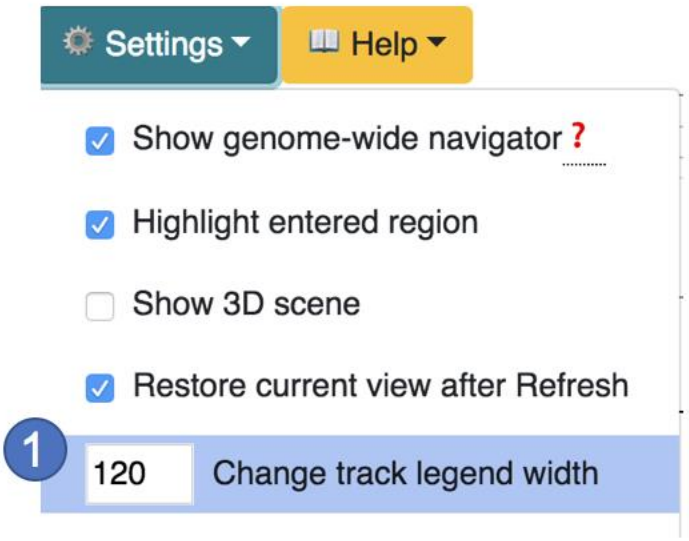


- Click on the **X** at the top-right of the window to get back to the browser.

3.4. Configuring the Epigenome Browser

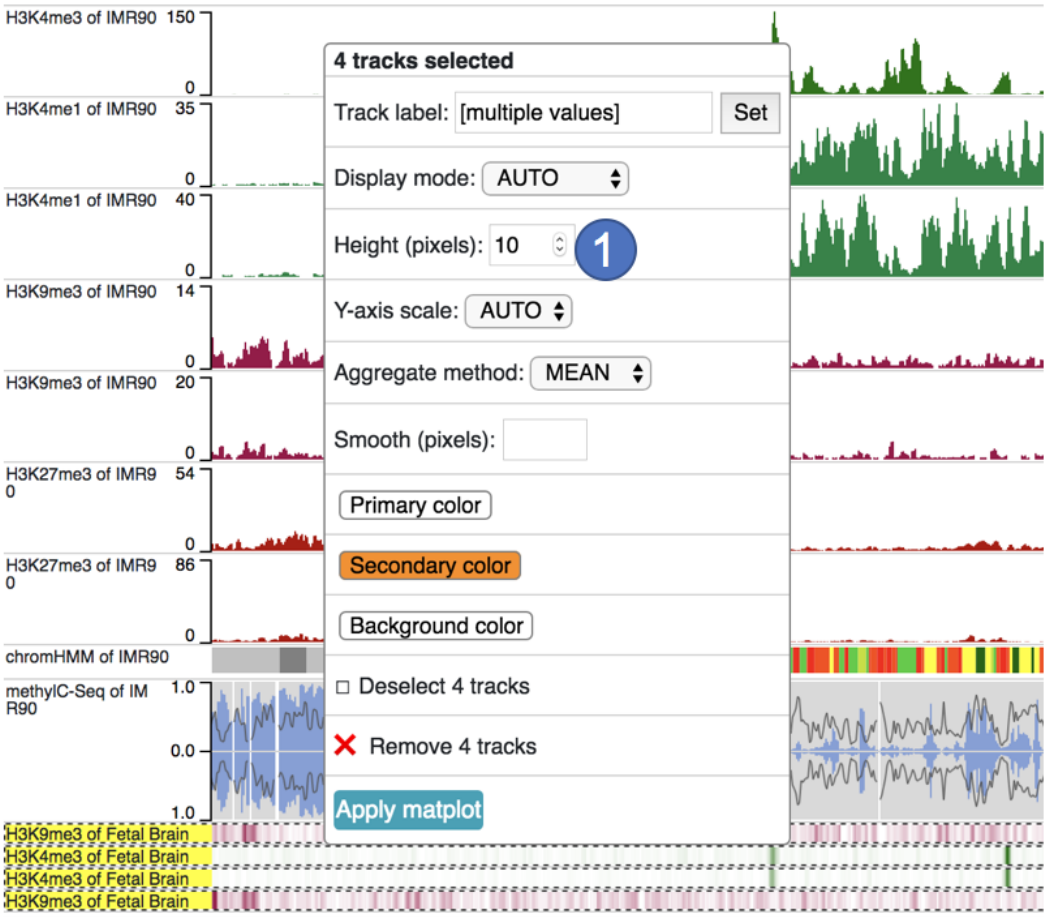
3.4.1. Change track label width

- The **width of the track names** can be changed using **Change track legend width** option under the Settings menu.



3.4.2. Changing the height of tracks

- To select multiple tracks, hold down the **shift-key** and click on the names of the 4 **Fetal Brain** tracks that was just added. Right-click on the yellow-highlighted track names. This will bring the configuration window.
- Change the value in **Height** option to increase/decrease the height of the tracks to your preference.



- Click outside this floating box to get back to the browser.
- To de-select the 4 tracks, right-click on the yellow-highlighted track names and select **Deselect 4 tracks**.

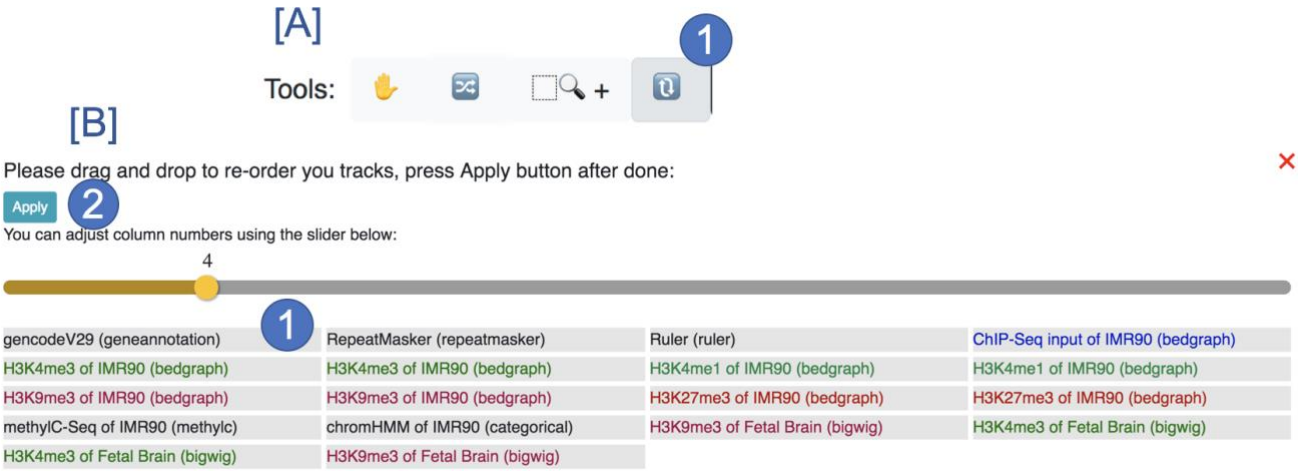
3.4.3. Re-ordering tracks

3.4.3.1. Using the Re-order tool

- [A] Choose the **Re-order** tool.
- [B] Drag the methylC track to one track up.

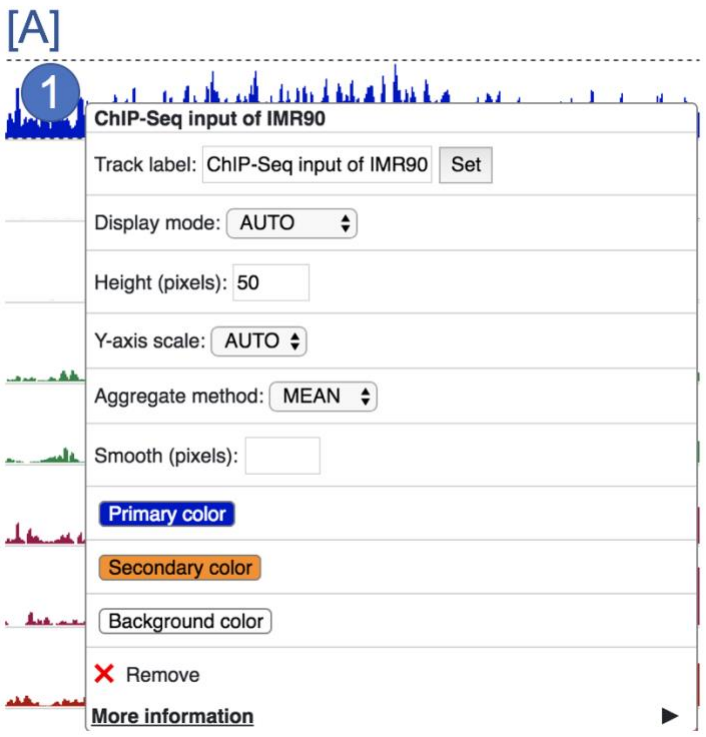
3.4.3.2. Using the Re-order Many app

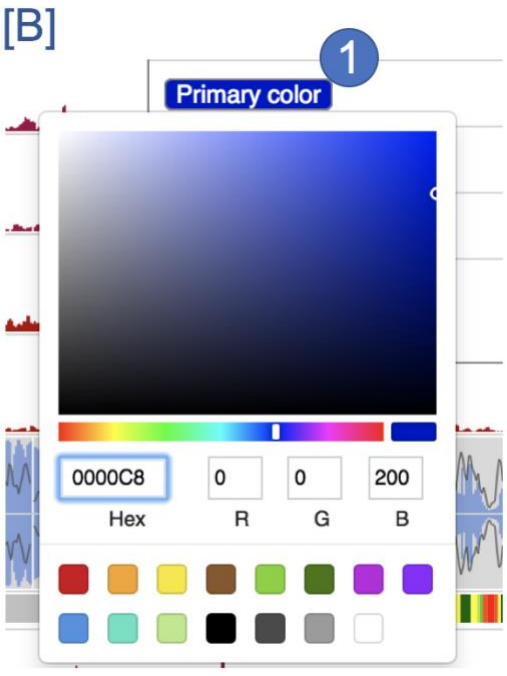
- [A] Choose the **Re-order Many** tool.
- [B] Drag the tracks with any order you like, press the **Apply** button when you are done.



3.4.4. Changing the color of tracks

- [A] Right Click on the **Chip-Seq input of IMR90** track.
- [B] In the **Configure** menu, click on the **Primary color** button and then select the color of your choice to change the color of the track.

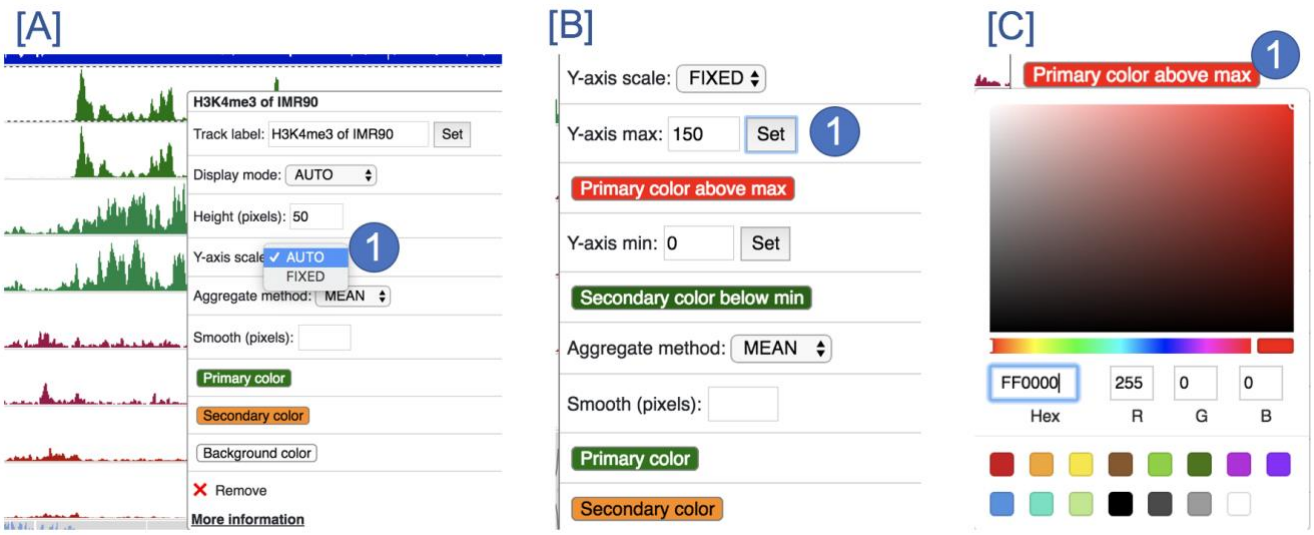




- Click outside this floating box to get back to the browser.

3.4.5. Changing the y-axis scale for tracks

- [A] Right-click on one H3K4me3 track.
- [B] Click on the **Y-axis scale** drop-down menu and then click **FIXED**. This will generate text boxes to enter the y-axis range. Enter **150** in the max text-box. Click on **Set** to change the y-axis scale.
- [C] Click on the button labeled **Primary color above max**. This will generate a floating window with color options; change the color to any color of your choice.

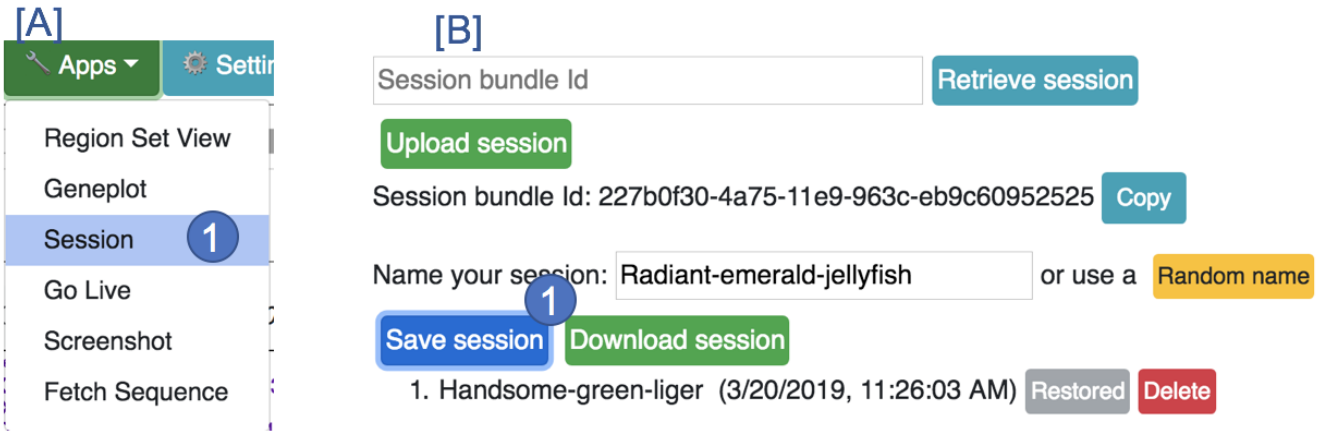


- Click anywhere outside the **Configure** box to get back to the browser.

4. Sessions: saving and retrieving browsing sessions

4.1. Saving sessions

- [A] Click on the **Apps** menu, and then select **Session**.
- [B] Click the **Save session** button.



4.2. Sharing links for collaboration

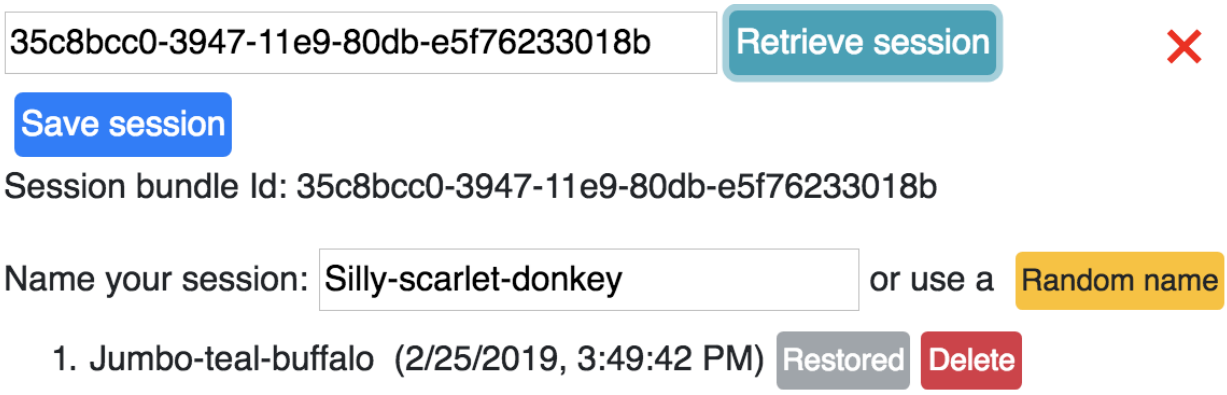
- Copy the session bundle id, and the session can be accessed using link:

  https://epigenomegateway.wustl.edu/browser/?bundle=session_bundle_id

4.3. Retrieving sessions

4.3.1. Using session bundle id

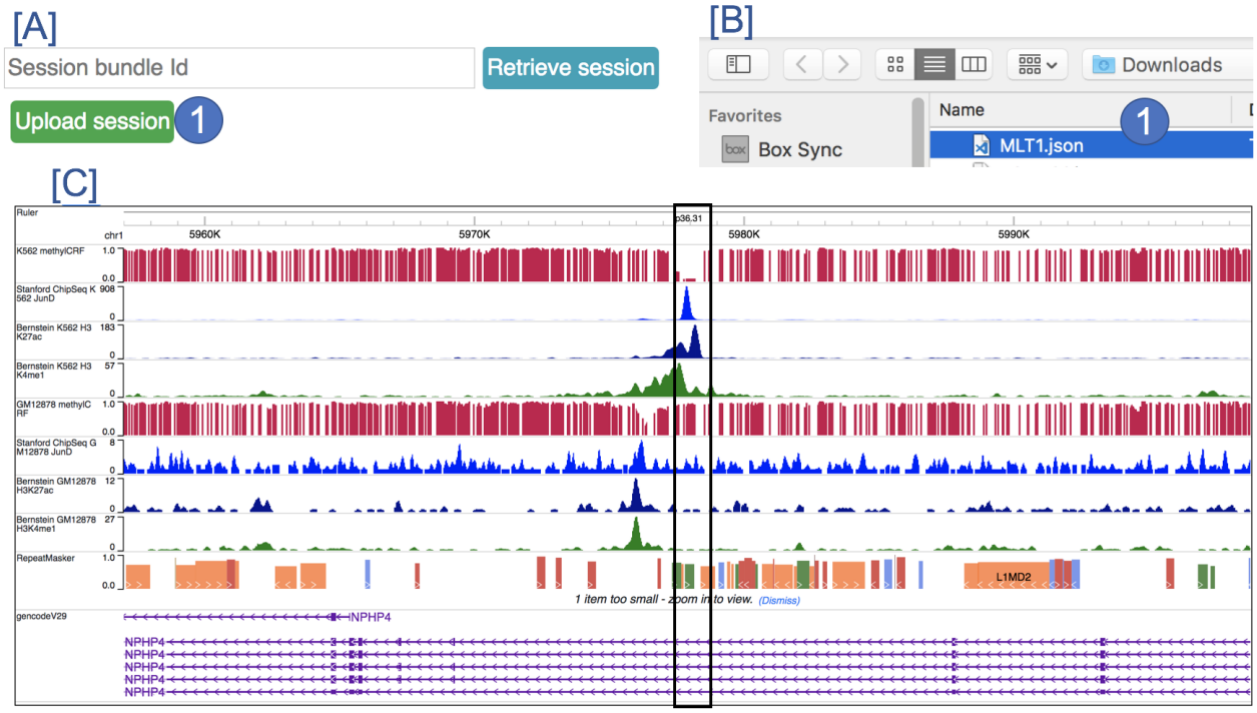
- Paste the session bundle id to the input box, and click the **Retrieve session** button.



4.3.2. Upload a session file

- Launch a new instance of the EpiGenome browser in a new window (URL: http://epigenomegateway.wustl.edu/browser).
- [A] Click the **Upload session** button.
- [B] Choose the **MLT1.json** file provided on the workshop website.

➢ The browser will load this session as shown in [C].

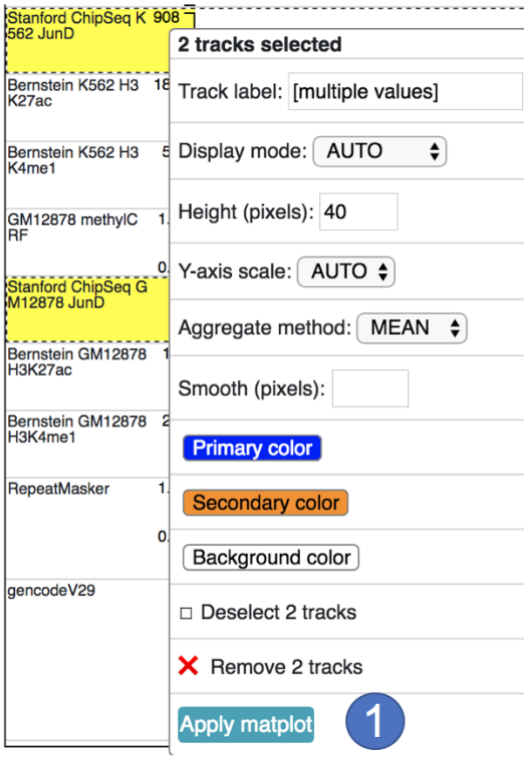


- In this session, we will be using ENCODE datasets to analyze the cell-type specificity of JunD binding on transposable elements.
- The **MLT1** session shows the specific binding of JunD to a transposable element, MLT1 in K562, specifically. This is corroborated by K562-specific hypomethylation of MLT1, and GM12878-specific hypermethylation.

5. Using Apps and Functions on the EpiGenome Browser

5.1. Matplot: Compare two or more data tracks by plotting datasets to one y-axis scale.

- Select the two **JunD ChipSeq** tracks and right-click on any of the tracks' names. Click on **Apply matplot**.



- This results in a track where the ChipSeq data for JunD binding in K562 and GM12878 share the same y-axis, and can be easily compared. It is evident from this view, that there is much more JunD binding in K562 compared with GM12878, on this MLT1 element.

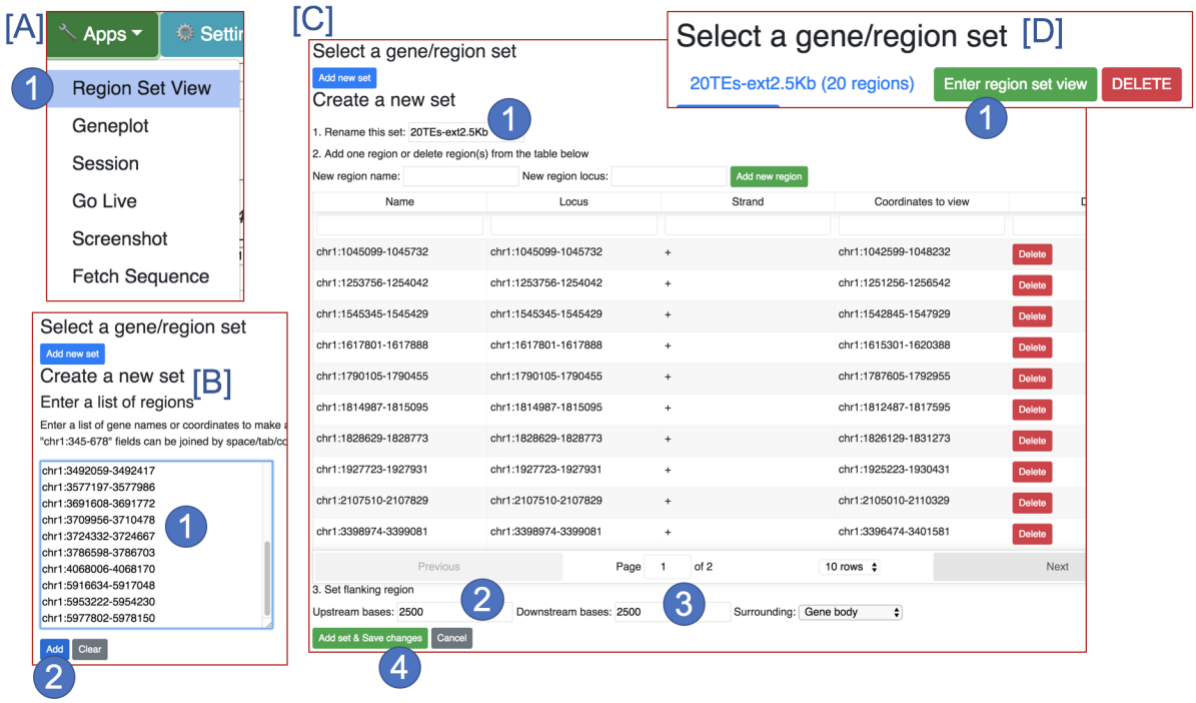
- To exit from the matplot view, right-click on the matplot track name and click on **Remove**.

5.2. Region-set: Visualizing multiple genomic regions in parallel.

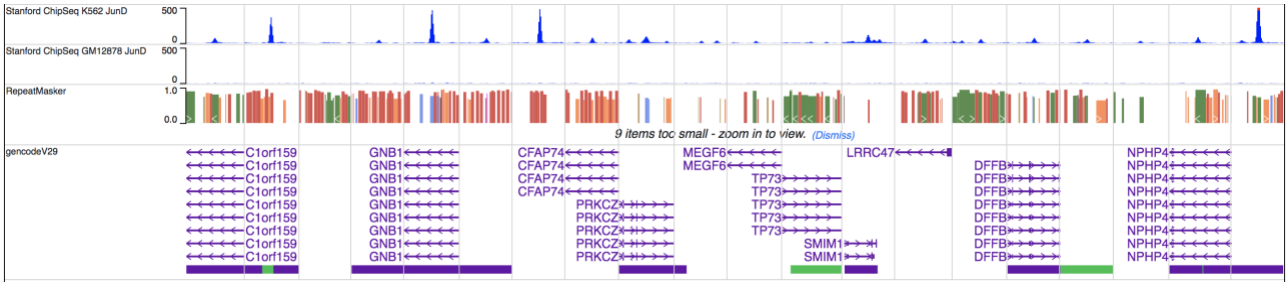
- [A] To view data on multiple genomic loci at a time, click on the **Apps** menu and then select **Region Set View**.
- [B] Clear the placeholder list, and input following 20 regions.

```
chr1:1045099-1045732
chr1:1253756-1254042
chr1:1545345-1545429
chr1:1617801-1617888
chr1:1790105-1790455
chr1:1814987-1815095
chr1:1828629-1828773
chr1:1927723-1927931
chr1:2107510-2107829
chr1:3398974-3399081
chr1:3492059-3492417
chr1:3577197-3577986
chr1:3691608-3691772
chr1:3709956-3710478
chr1:3724332-3724667
chr1:3786598-3786703
chr1:4068006-4068170
chr1:5916634-5917048
chr1:5953222-5954230
chr1:5977802-5978150
```

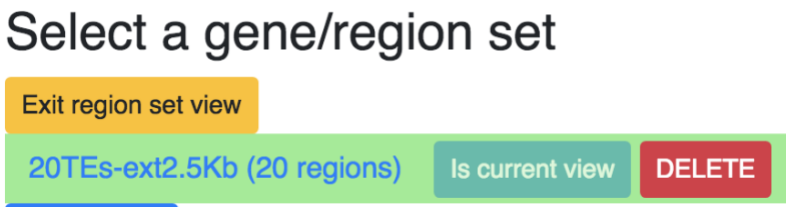
- [C] Rename this set to **20TEs-ext2.5Kb**, set flanking region to upstream 2500bp and downstream 2500bp, click the **Add set & Save changes** button.
- [D] To view this region set, click the button **Enter region set view**.



- This results in a tiled-view of all the TEs arranged beside each other. By setting the same y-axis scale for the two JunD ChipSeq tracks, the K562-specificity of JunD binding on these TEs is evident.



➢ To exit the region set view, go back the region set interface (Apps -> Region Set View), and click the yellow **Exit region set** view button.



5.3. **Gene plot:** Summarize the data distribution across multiple regions, by splitting the regions into bins.

➢ [A] Click on the **Apps** menu. choose **Geneplot** app.

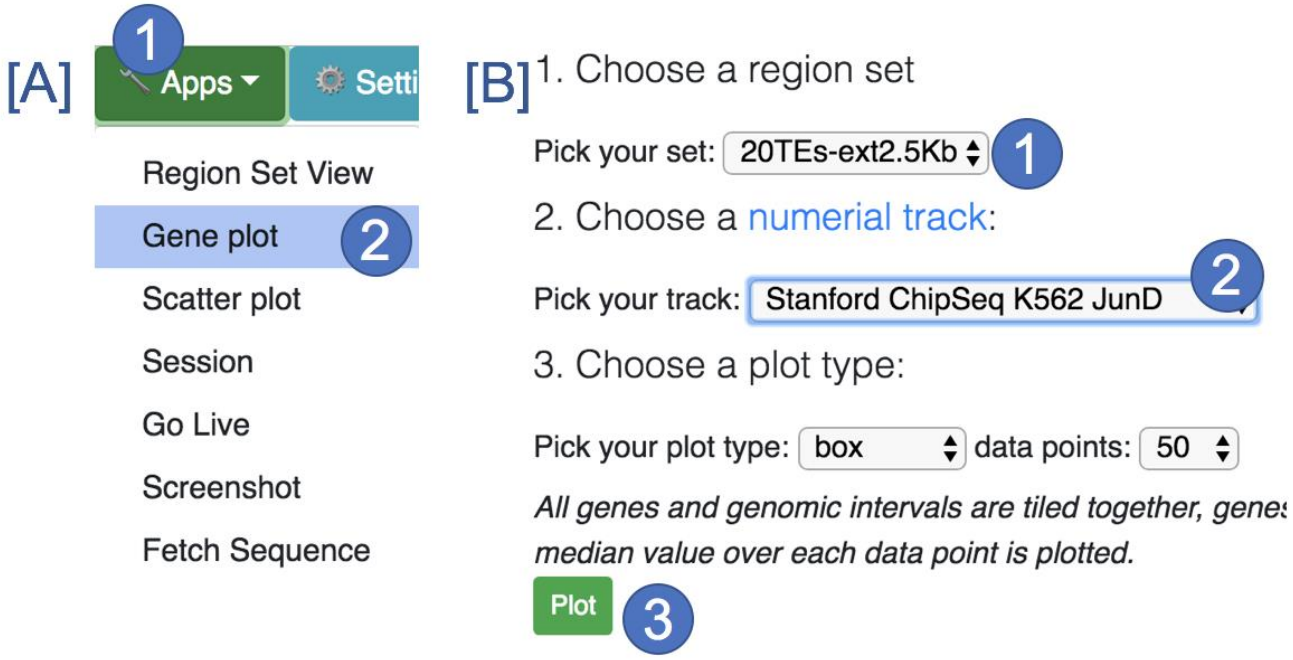
➢ [B] To make a gene plot, follow these steps:

  ➢ Click on the **Choose a region set** dropdown menu. Select the **20 TEs-ext2.5Kb** list of TEs for this analysis.

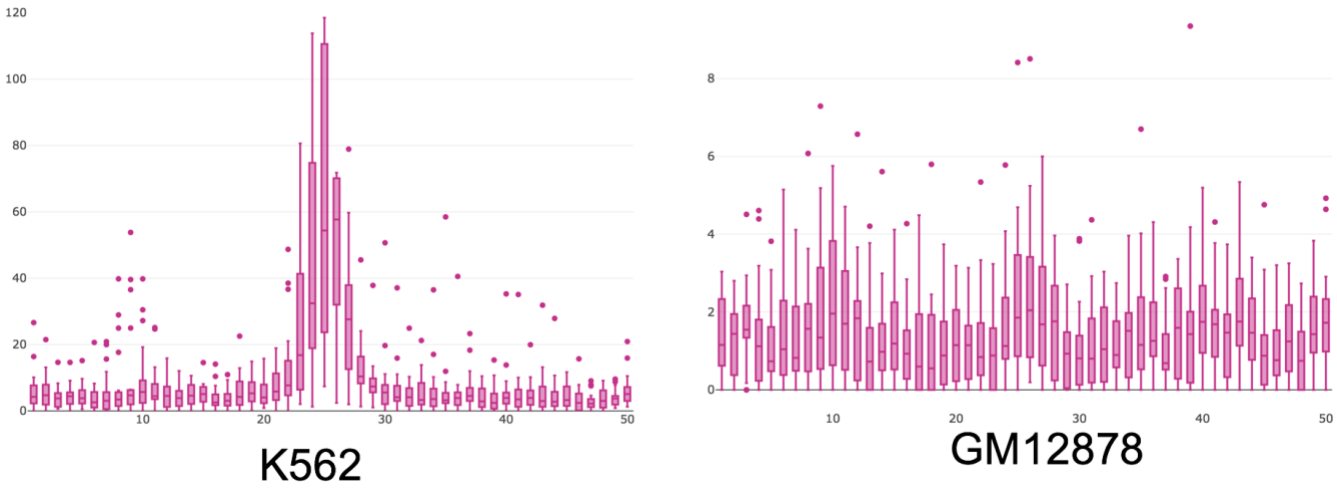
  ➢ Click on the **Choose a numerical track dropdown menu**, and then select **Stanford ChipSeq K562 JunD**.

  ➢ Click the **Plot** button.

  ➢ The same process can be repeated for the **Stanford ChipSeq GM12878 JunD**.



- Comparing the two gene-plots reveals that the highest ChIP-seq signal for JunD binding in K562 is on the TEs (data points 23-29 on the x-axis). The ChIP-seq signal for JunD binding in GM12878 on TEs is comparable to the flanking region (data points 1-23 and 29-50).



Click the top and bottom number of y-axis can change/edit the y-axis scale. When mouse over

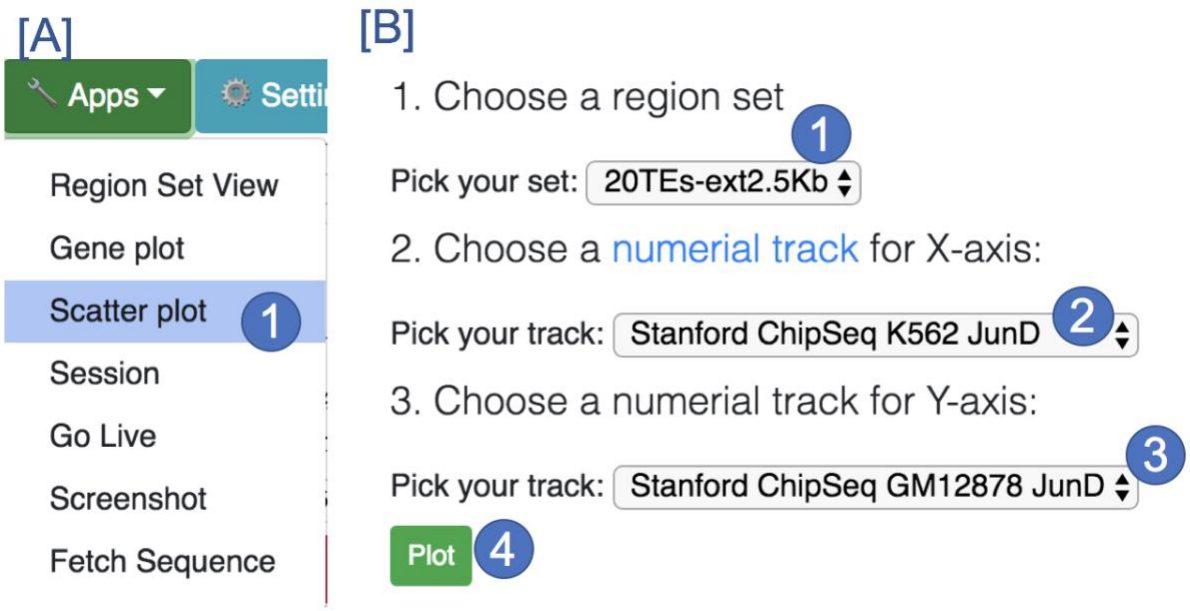
the figure, click **Autoscale** button to reset the scale.

5.4. Scatter plot: Correlate different datasets, across multiple genomic regions.

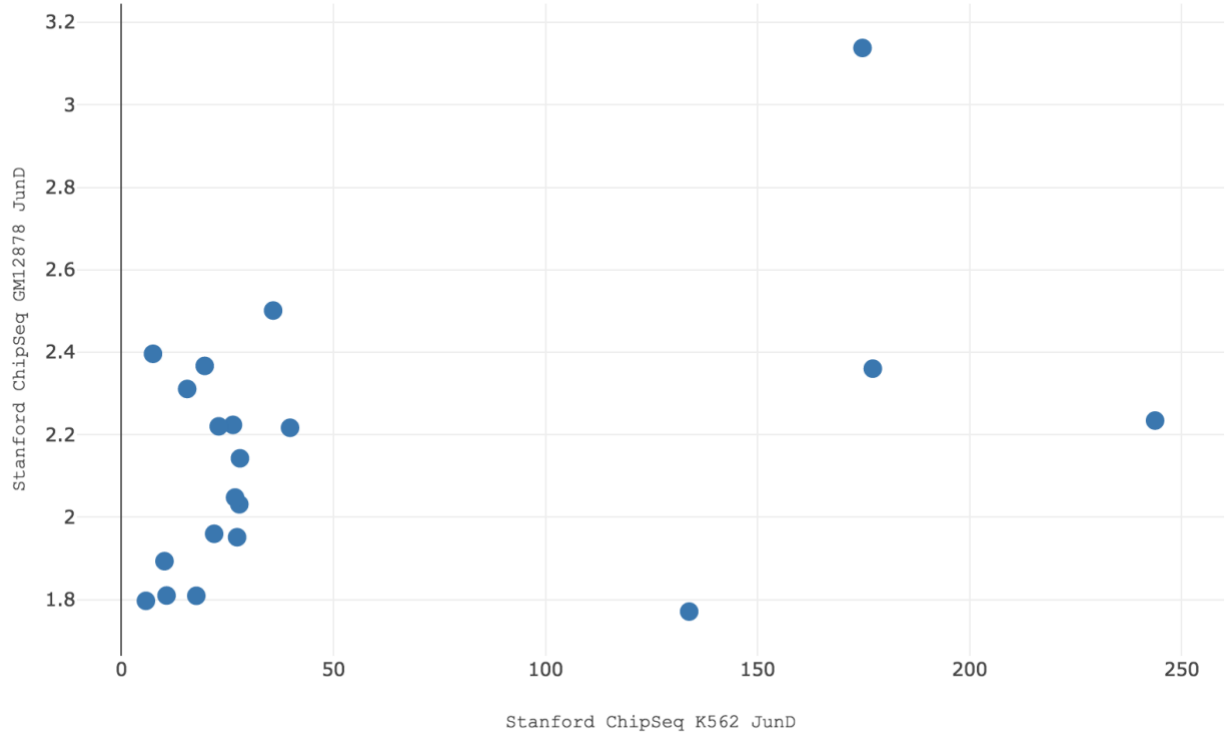
[A] Click on Apps -> Scatter plot

[B] Choose region set, 2 tracks for X- and Y-axis, repspectively.

[C] Click the Plot button.



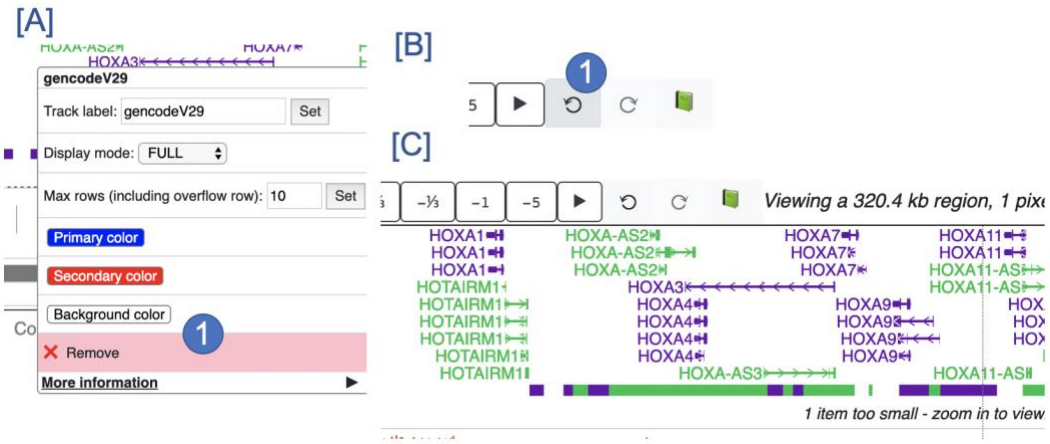
Below is the scatter plot generated:



5.5. Undo/Redo/History

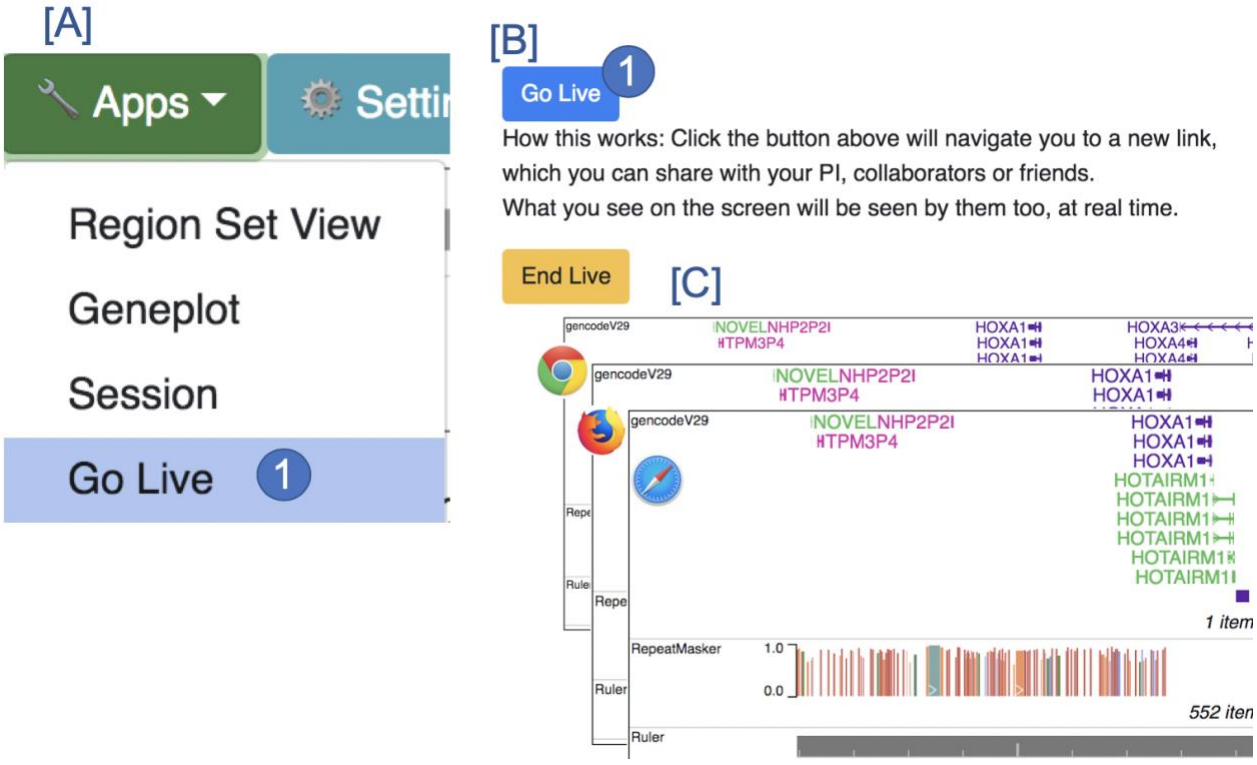
The browser now enables undo/redo and history function. Launch a new session with hg19 genome loaded.

- [A] Right click the gencodeV29 track, choose Remove
- [B] Click the Undo button.
- [C] The Removed gencode track is added back.



5.6. Living Browsing

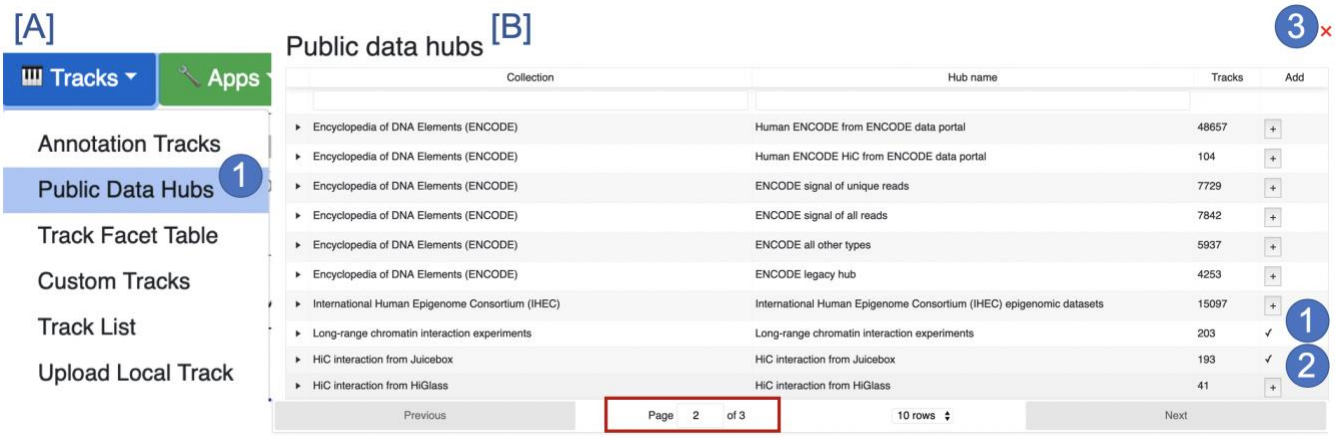
- [A] Click Apps -> Go Live
- [B]Click the button Go Live.
- [C] Send the generated link to someone else. Ask s/he to open the link.
- Operate on the browser, s/he could see the same browser view as you see. Or ask her/him to operate, you can see the same view as her/him.



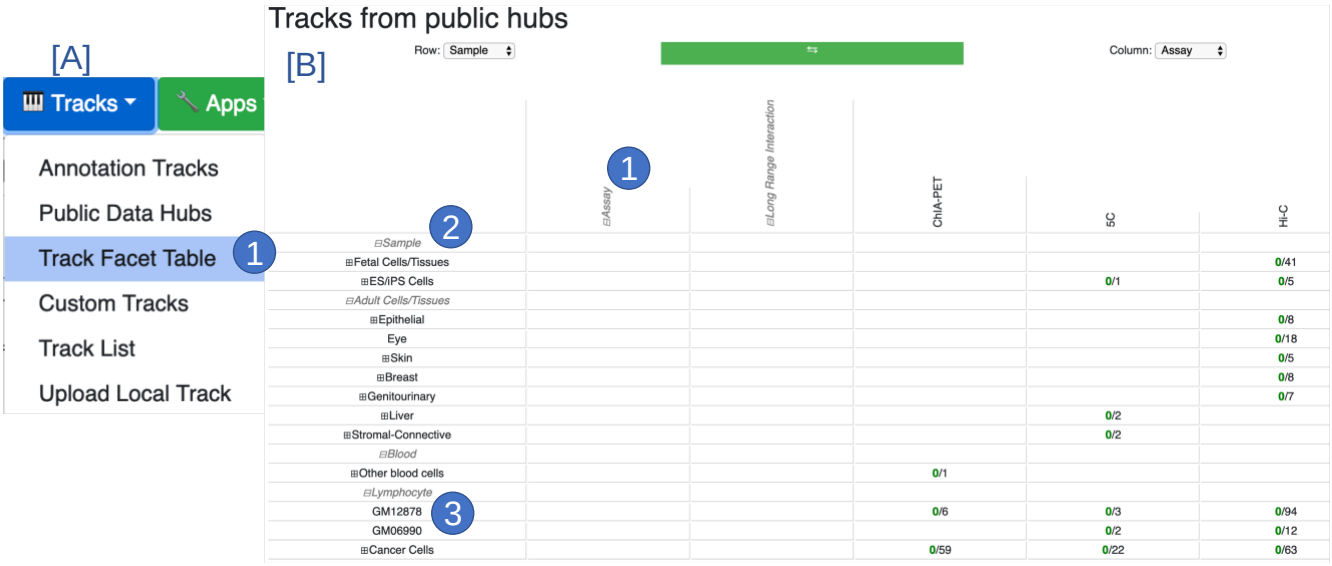
6. Exploring chromatin interaction data using the EpiGenome Browser

Let's load some chromatin interaction data to the browser.

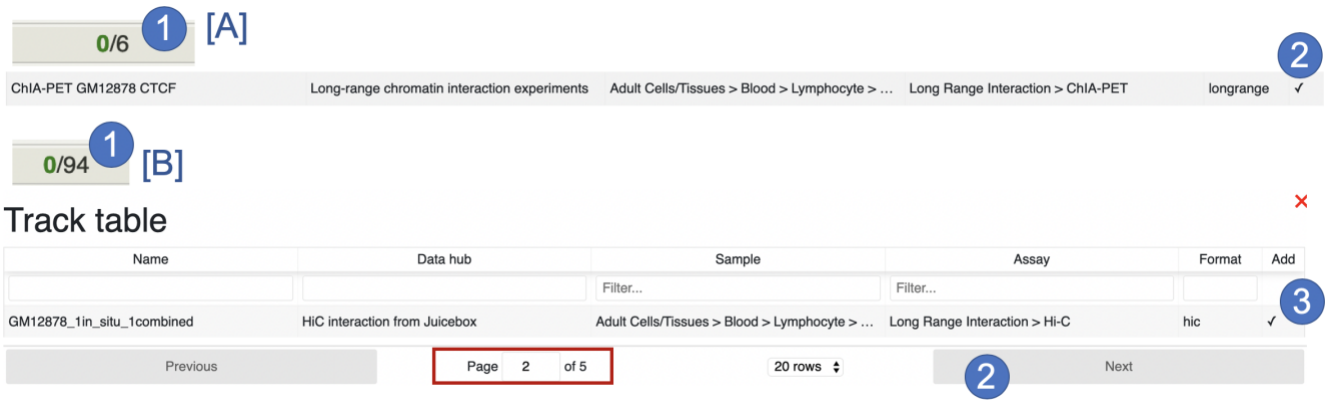
- [A] Click **Tracks -> Public Data Hubs**.
- [B] Go to 2nd page of data hub list, load **Long-range chromatin interaction experiments** and **HiC interaction from Juicebox** hub.
- Click on the **X** at the top-right of the window to get back to the browser.



- [A] Click Tracks -> Track Facet Table.
- [B] Expand Sample by clicking the Sample term, Find GM12878 Under Adult Cells/Tissues -> Blood -> Lymphocyte, expand the Assay column as well.

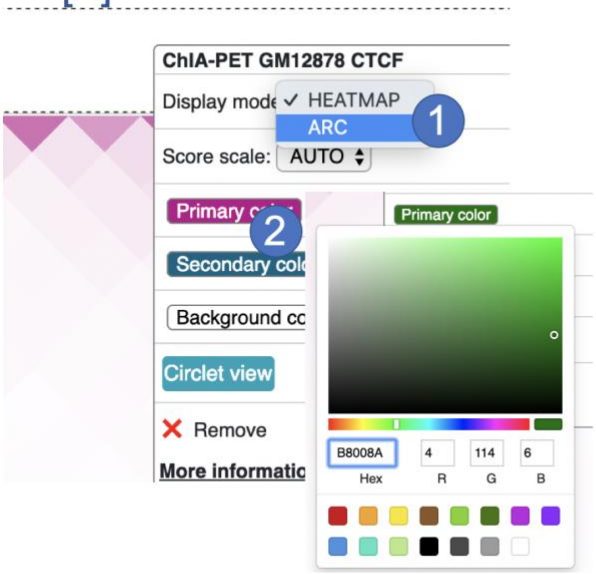


- Click the 0/6 cell in ChiA-PET column and GM12878 row, load the track named **ChIA-PET GM12878 CTCF**.
- Click the 0/94 cell in HiC column and GM12878 row, load the first track in 2nd page of track list, named **GM12878_1in_situ_1combined**.
- Click on the **X** at the top-right of the window to get back to the browser.

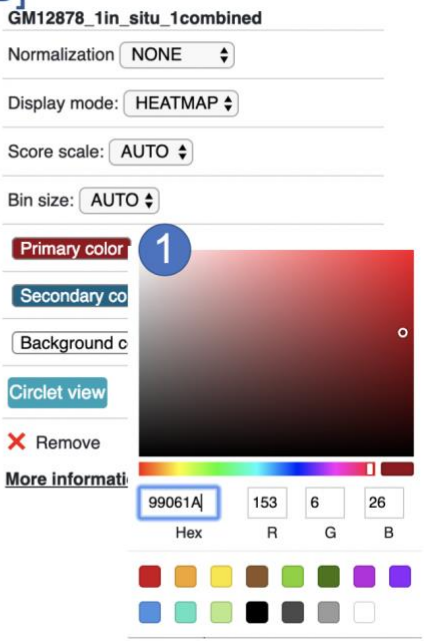


- [A] Right click the ChIA-PET track, change Display mode to ARC, and Click the Primary color button to change color.
- [B] Right click the hic track, click the Primary color button to change color.

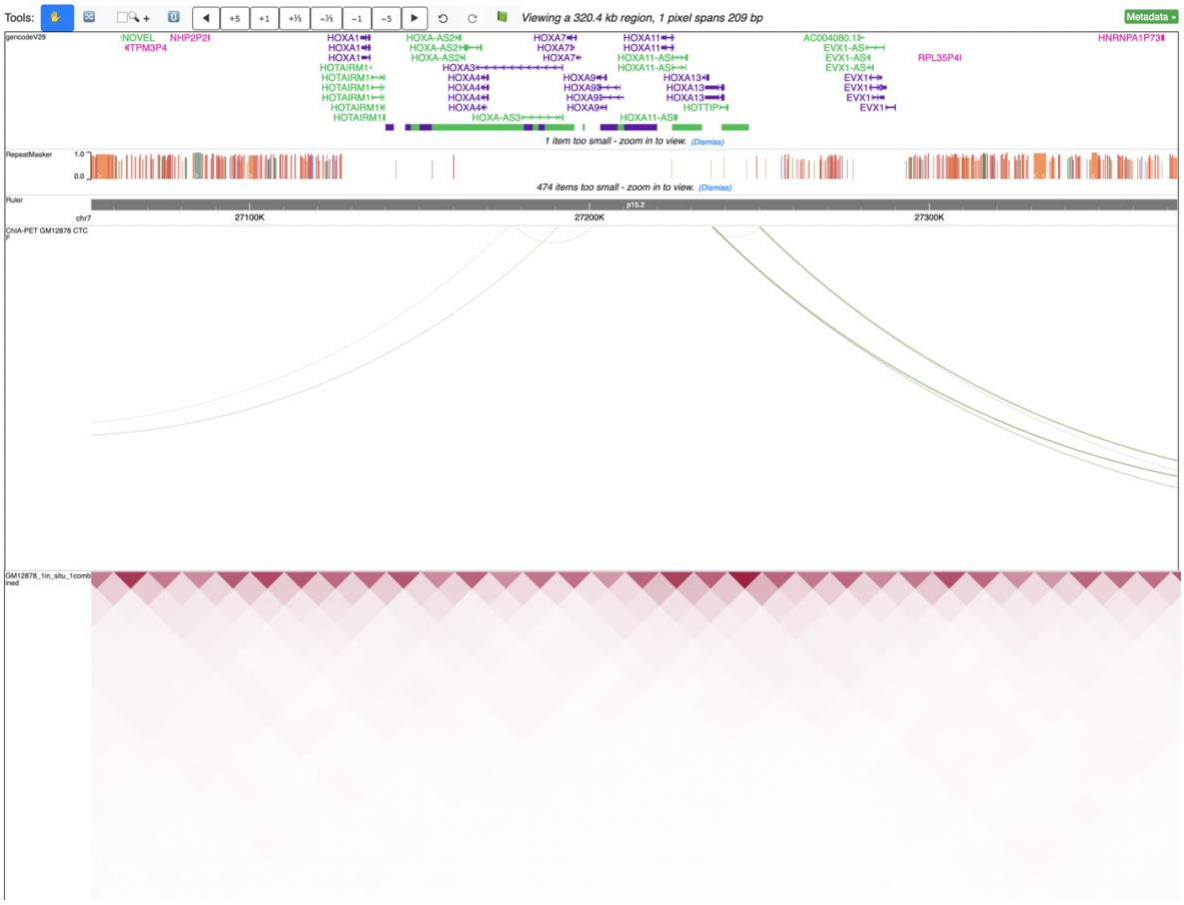
[A]



[B]

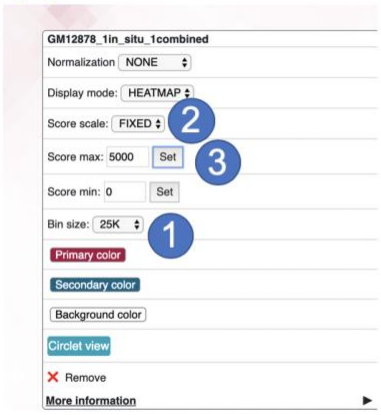


Now the browser view should look like this:



Click the -5 button to zoom out the view region.
[A] Right click on the hic track, change Bin size to 25K, and Score Scale to FIXED, and enter Score max to 5000, click the Set button.
[B] The updated hic track view after change bin size and scale.

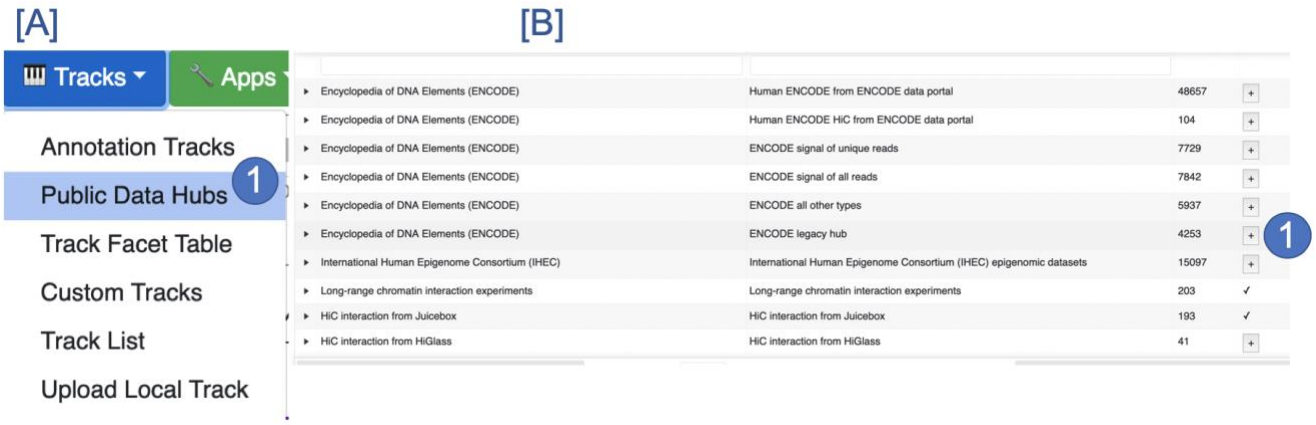
[A]



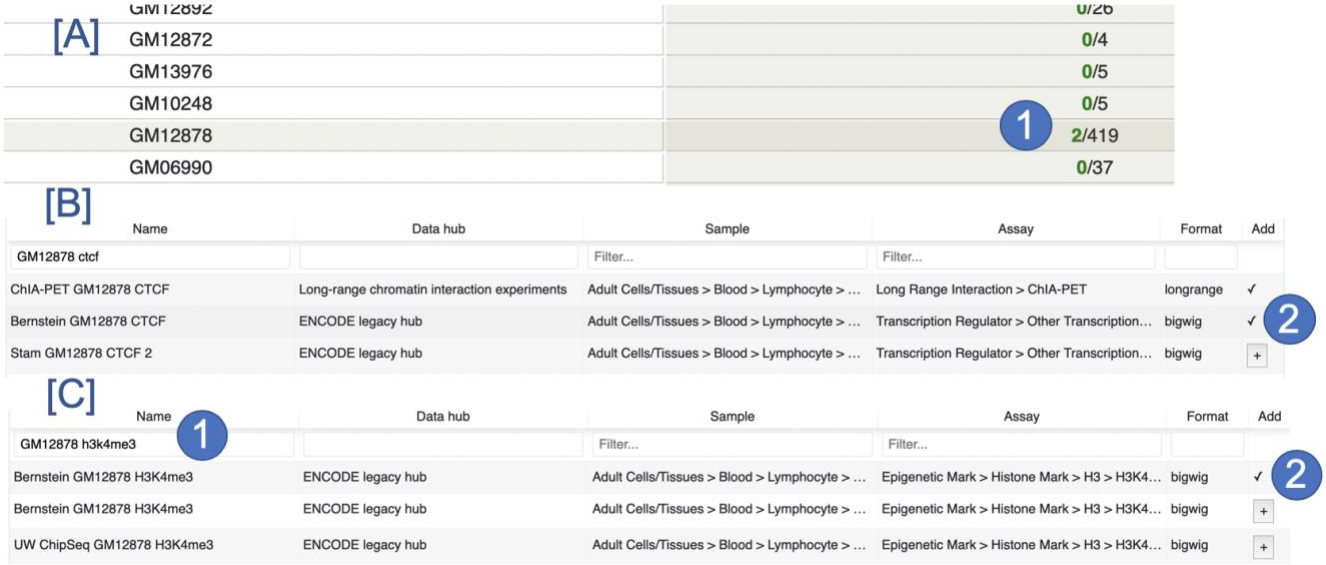
[B]

Let's load more epigenetic data.

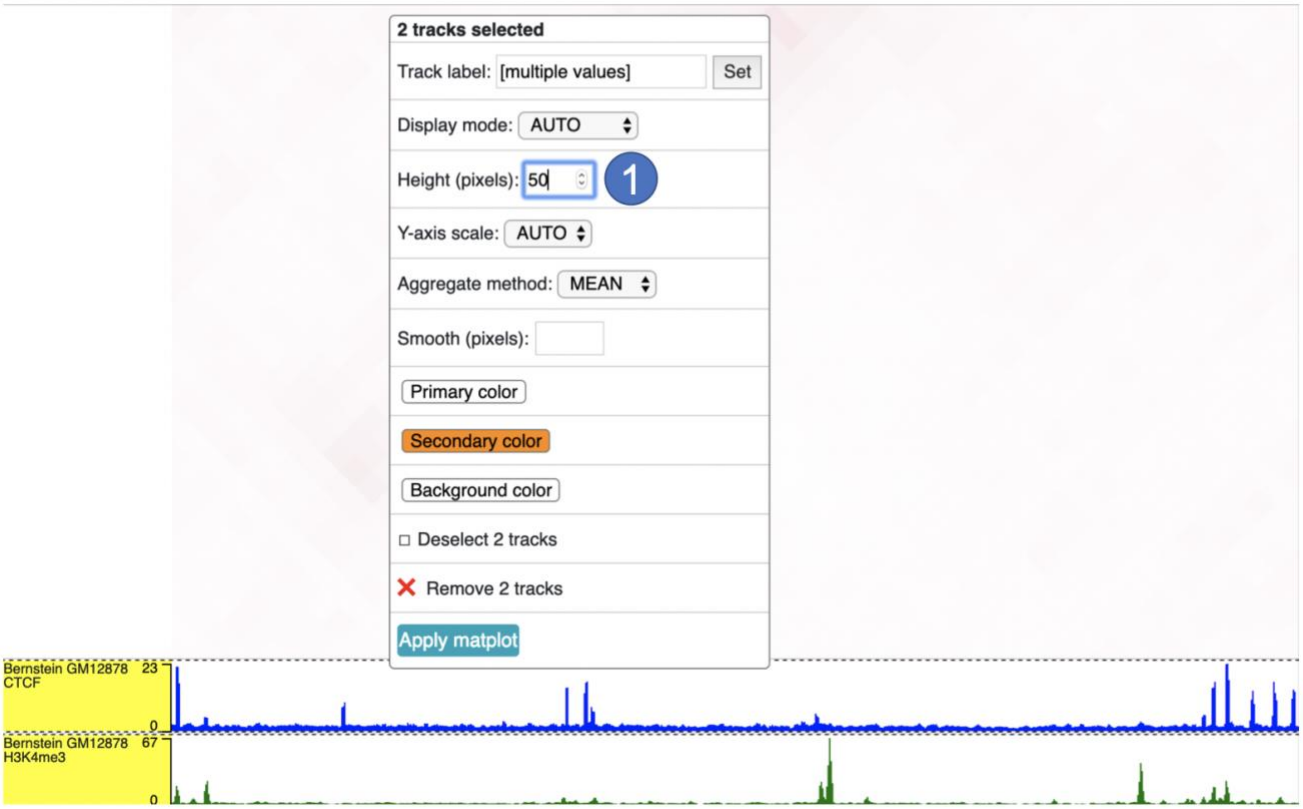
- [A] Click Tracks -> Public Data Hubs.
- [B] Load the ENCODE legacy hub.
- Click on the **X** at the top-right of the window to get back to the browser.



[A] Find GM12878 from the sample metadata. Click the cell labeled 2/419.
[B] Type ‘GM12878 ctcf’ in name search box, load the track Bernstein GM12878 CTCF.
[C] Type ‘GM12878 h3k4me3’ in name search box, load the track Bernstein GM12878 H3K4me3.



Hold Shift key and select the 2 newly added tracks, and right click to change height to 50px.



More information:

- For more documentation and tutorials, visit: http://epigenomegateway.wustl.edu/support/index.html
- Contacting WashU Epigenome Browser:
  - Google groups: http://bit.ly/egGroup
  - Facebook: http://bit.ly/2SZpvpz
  - Slack channel: https://bit.ly/2T1OKmP
  - Twitter: @wuepgg