WASHU EPIGENOME BROWSER

Keystone Symposium on DNA and RNA Methylation

January 23rd, 2018

Fairmont Hotel Vancouver, Vancouver, British Columbia, Canada

Presenter: Renee Sears and Josh Jang

Tutorial Overview:

WashU EpiGenome Browser Tutorial

WashU EpiGenome Browser:
http://epigenomegateway.wustl.edu/browser

Roadmap EpiGenome Browser:
http://epigenomegateway.wustl.edu/browser/roadmap/

0. Tutorial key

- 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 ②

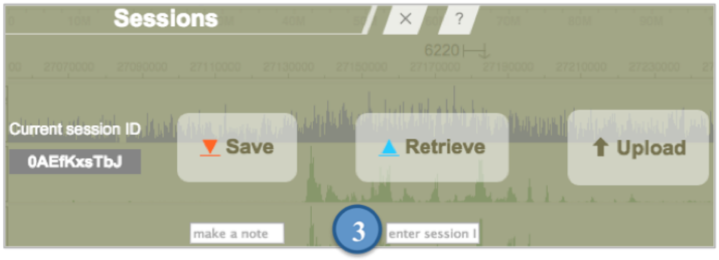
0.1 How to catch up

0.1.1 WashU EpiGenome Browser tutorial

- To catch up with any section of the *WashU EpiGenome Browser tutorial*, use the provided **session ID** and the **session status**, listed at the end of that section. For this, follow these steps:
  - ➢ [A] Click the **Apps** button to display the Apps menu. Then click **Session**.
  - ➢ [B] Under the **Retrieve** button, enter the **session ID** and then click the **Retrieve** button.
  - ➢ This will generate a list of **session statuses**. Click the session of interest.

[A]

[B]



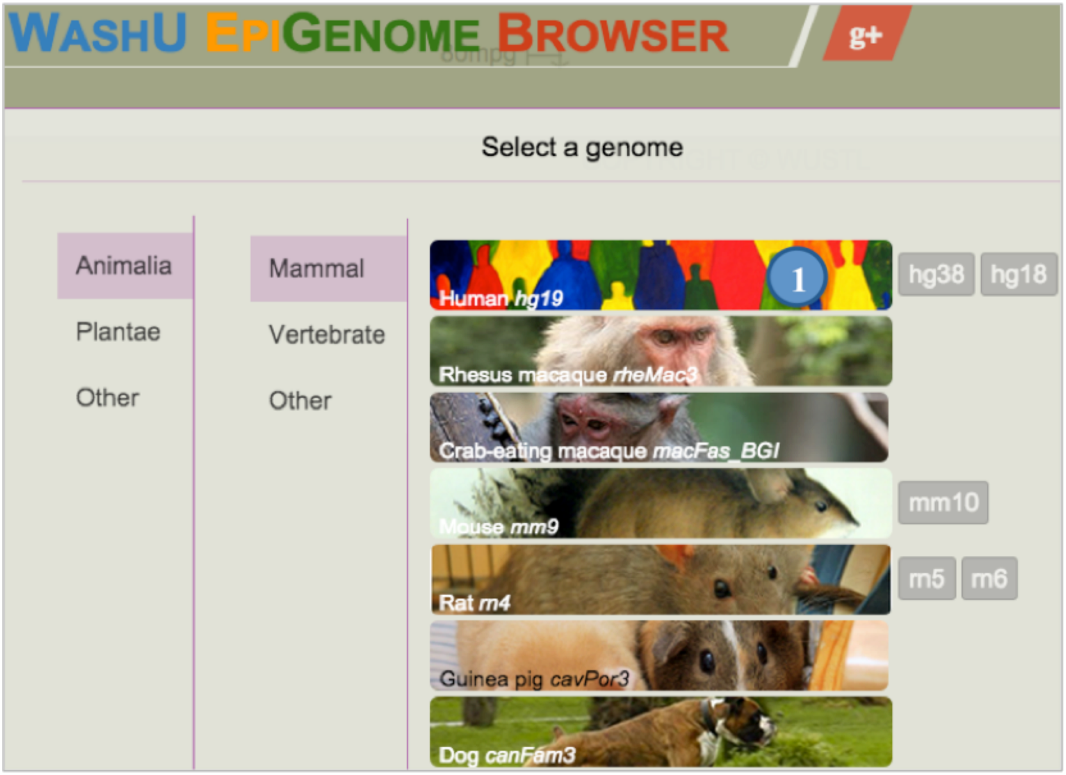
1. Getting started with the EpiGenome Browser

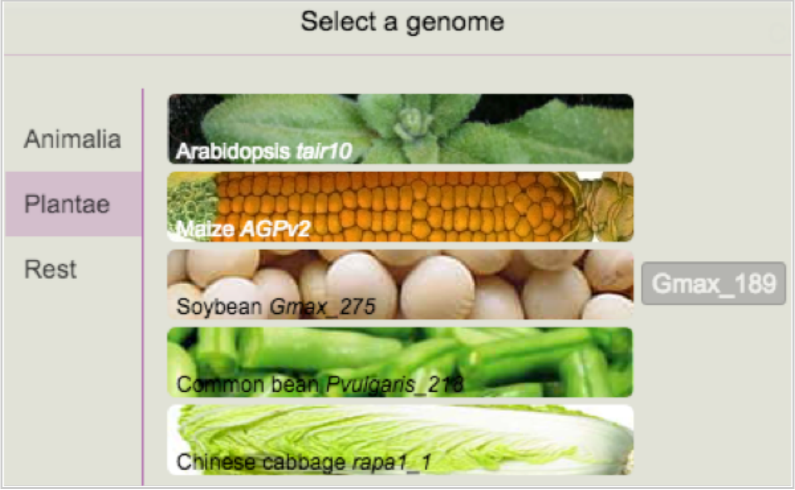
1.1. Loading the browser

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

1.2. Selecting the genome assembly of interest

- ➢ For the purpose of this tutorial, please select **Human hg19**.
- ➢ However, you can see that many other genomes are also available.

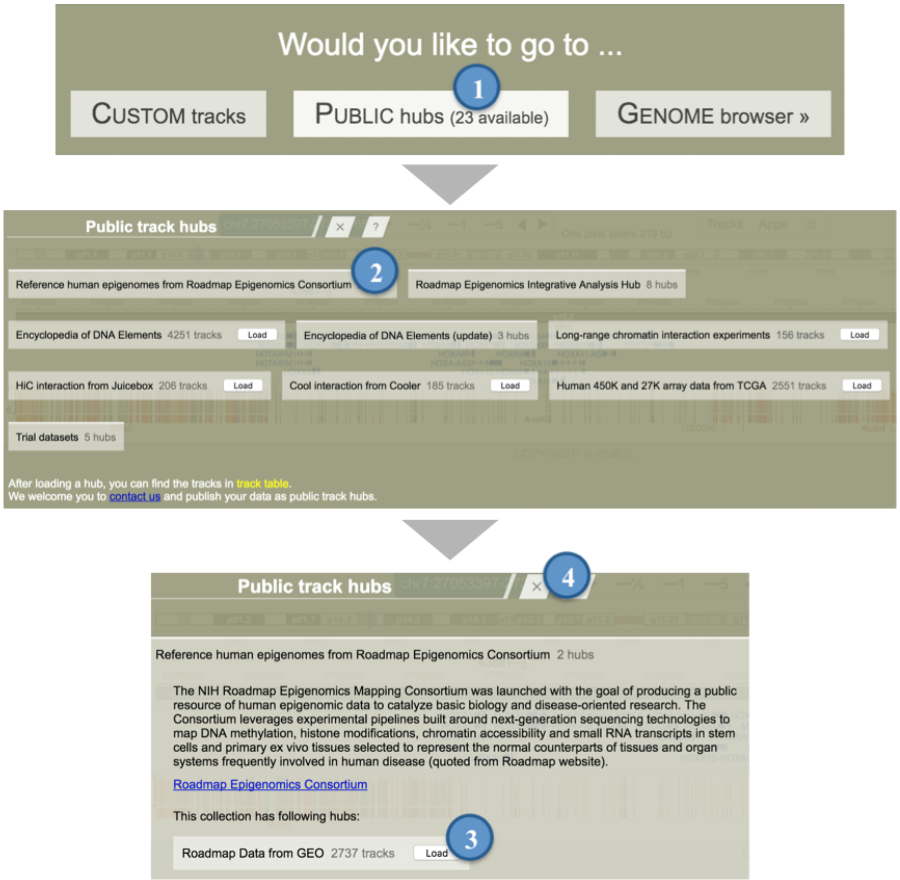




2. Loading data on the EpiGenome Browser

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

- [A] Click the **PUBLIC hubs (23 available)** button to view all the available public datasets on the EpiGenome Browser.
- [B] Click the **Reference human epigenomes from Roadmap Epigenomics Consortium (2 hubs)** button.
- [C] This will generate a list of available data hubs. Click the **Load** button on Roadmap Data from GEO (2737 tracks) box. Once the datasets are loaded, exit the data hub menu by clicking the **X** at the top-right of the panel or pressing **Esc**.



After loading a hub, you can find the tracks in track table.
We welcome you to contact us and publish your data as public track hubs.

The NIH Roadmap Epigenomics Mapping Consortium was launched with the goal of producing a public resource of human epigenomic data to catalyze basic biology and disease-oriented research. The Consortium leverages experimental pipelines built around next-generation sequencing technologies to map DNA methylation, histone modifications, chromatin accessibility and small RNA transcripts in stem cells and primary ex vivo tissues selected to represent the normal counterparts of tissues and organ systems frequently involved in human disease (quoted from Roadmap website).

This collection has following hubs:

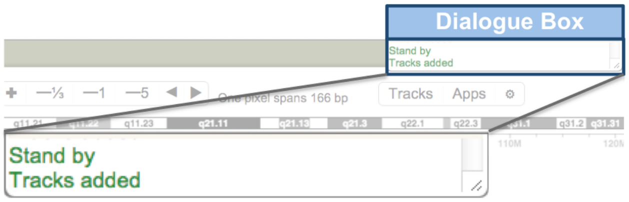
3. Navigating the EpiGenome Browser

3.1. Layout of the EpiGenome Browser

Tracks added
Stand by

3.1.1. Dialogue box

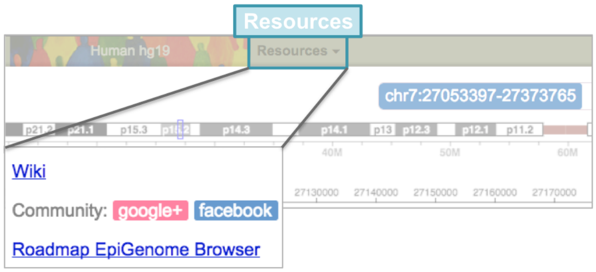
➢ This box logs all the actions performed by the EpiGenome browser. Any errors the browser encounters are displayed.



Stand by
Tracks added

3.1.2. Resources

➢ This window contains links to the documentation on the browser (**Wiki**); social-media discussion forums (on **Google+** and **Facebook**) for the browser; and the **Roadmap EpiGenome Browser**, a companion browser for visualizing and analyzing data from the Roadmap Epigenomics Consortium.



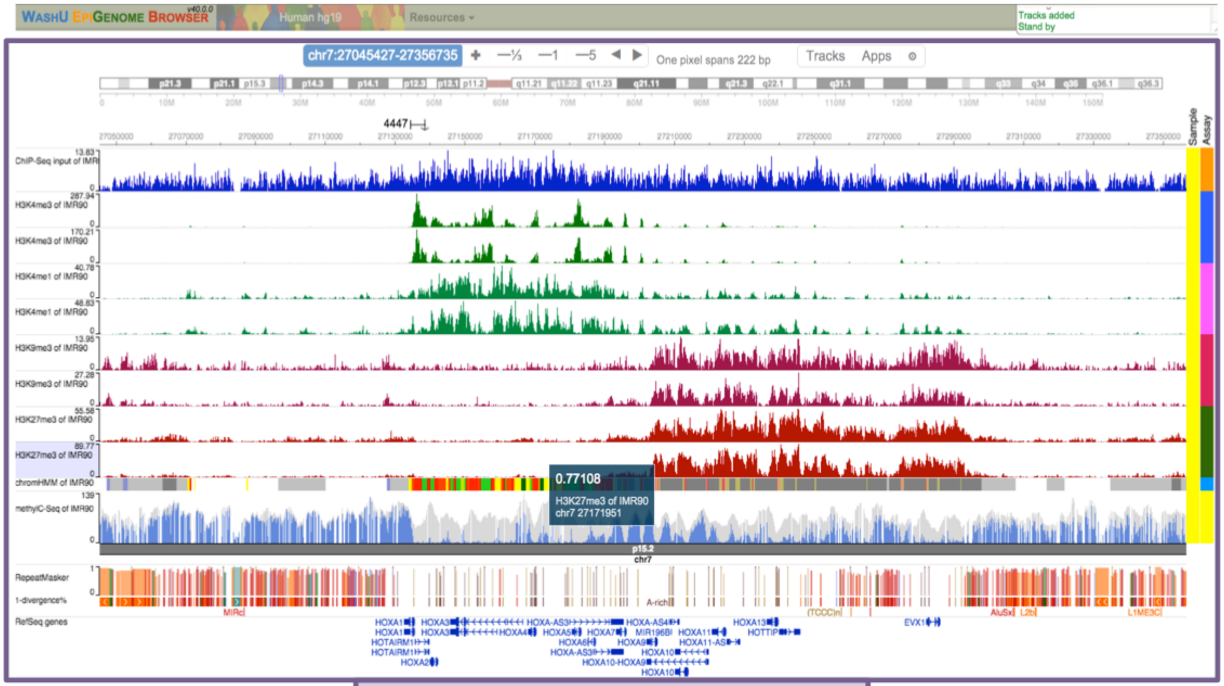
Wiki

Community: google+ facebook

Roadmap EpiGenome Browser

3.1.3. WashU EpiGenome Browser

- ➢ This region contains the data tracks aligned to the genome along with reference genes and a chromosome ideogram. Additional public track hubs, annotation tracks, and custom tracks can be added by the user.

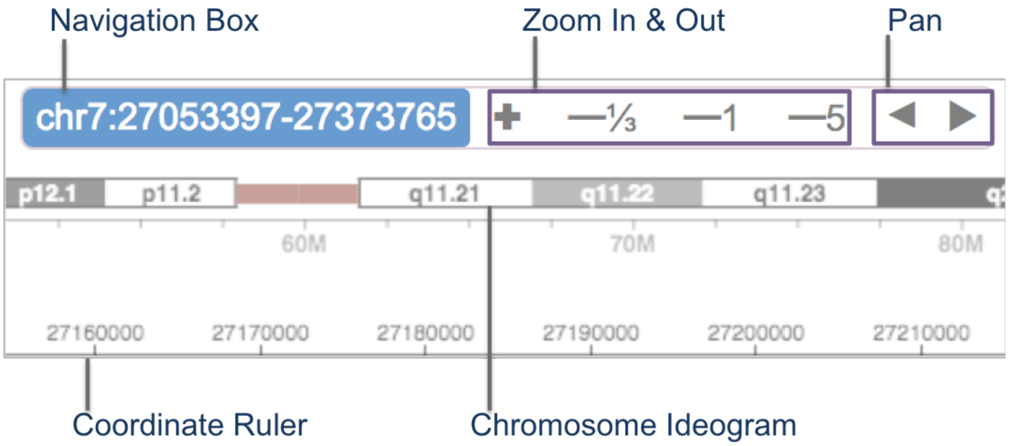


WashU EpiGenome Browser

3.2. Navigating the genome on the EpiGenome Browser

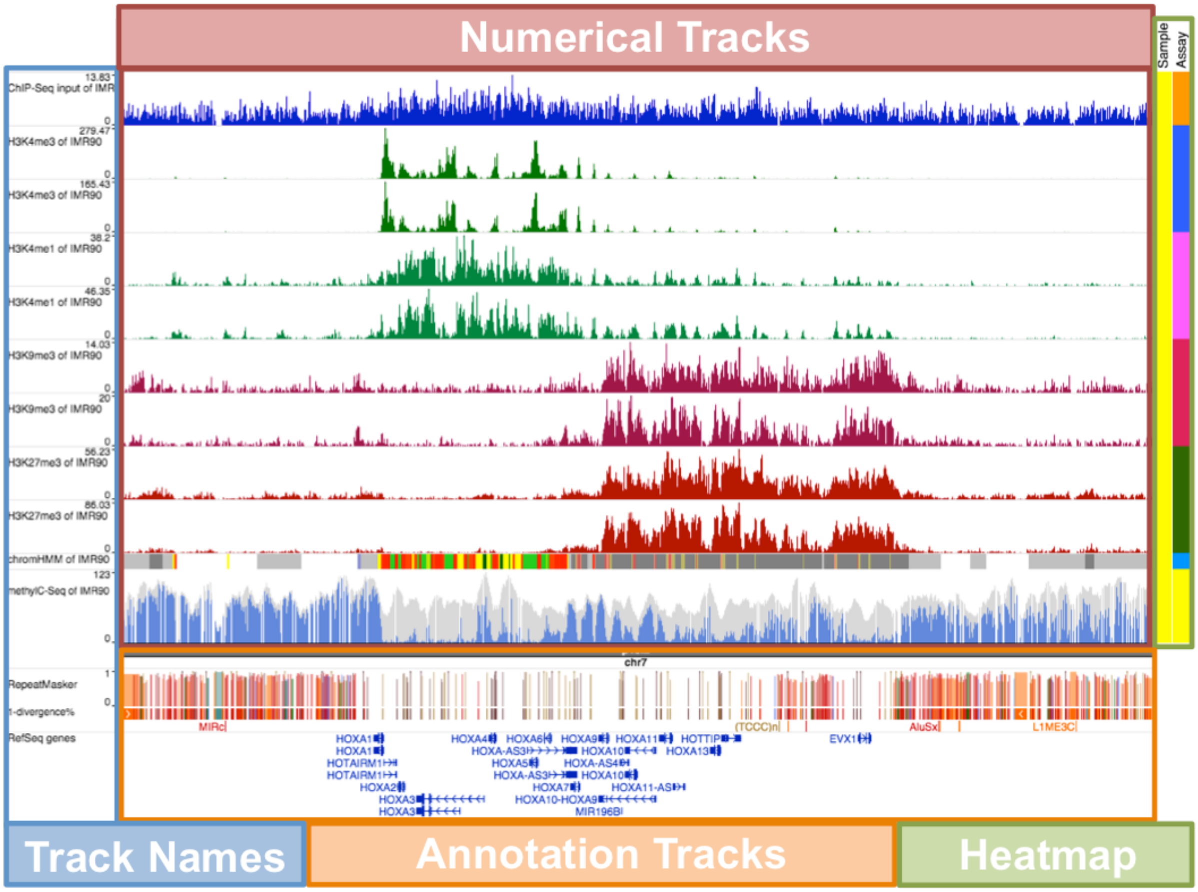
3.2.1. Genome navigation controls

- ➢ The controls enable exploration across the genome using the **zoom in**, **zoom out**, and **pan** buttons.
- ➢ Alternatively, click the **navigation box** to enter the genomic coordinates of a region of interest or a gene name. Click **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.
- **Track names** for each track are listed on the left-hand side of the track.
- On the right-hand side of the tracks is a **heatmap**, which represents the **metadata** for the tracks. Metadata are terms used for annotating tracks with experimental and sample information.
- Below the numerical tracks are **annotation tracks**, including annotations of genes and transposable elements.



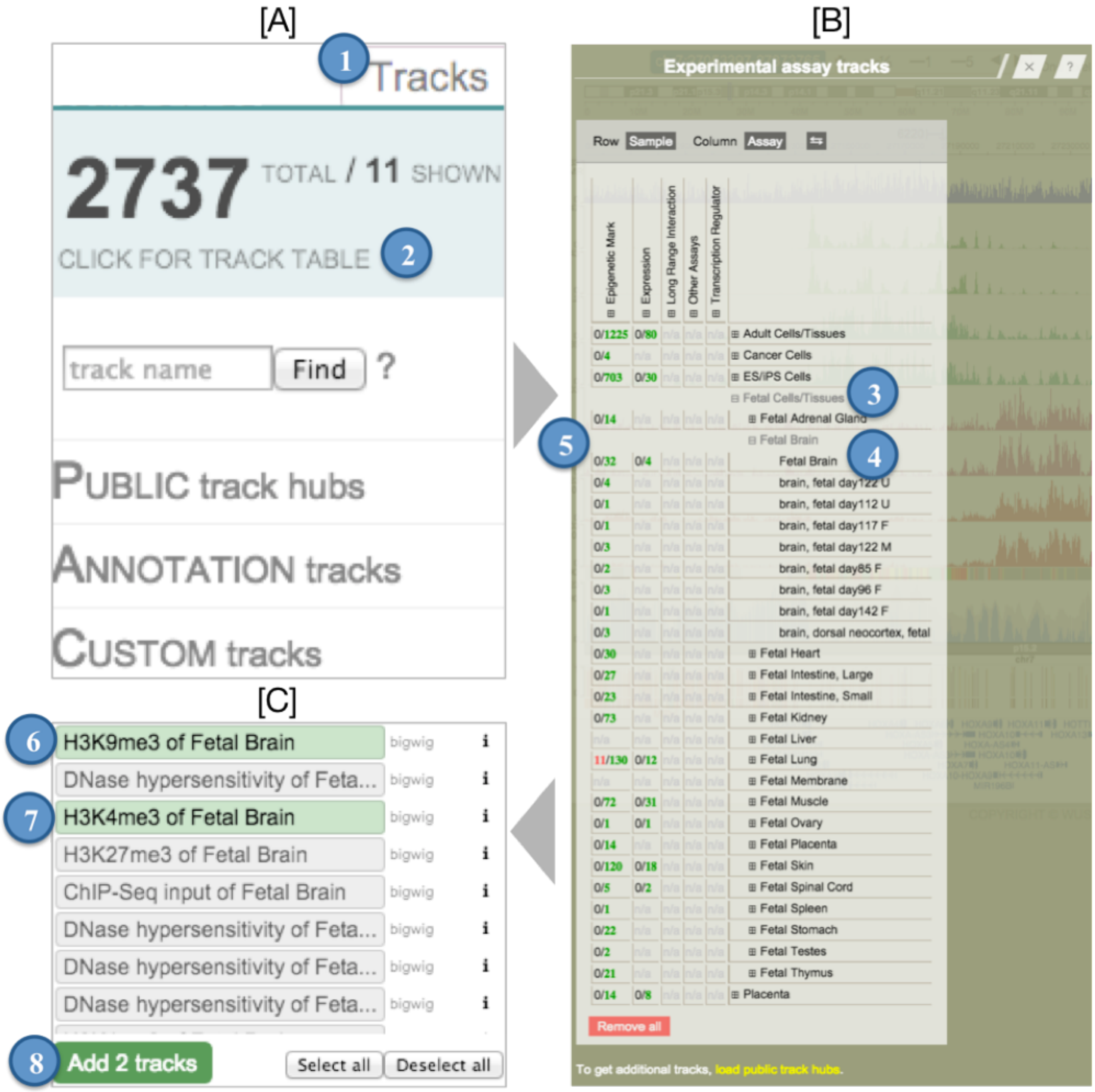
3.3. Adding datasets on the EpiGenome Browser

3.3.1. Searching for datasets using the facet table:

Searching for datasets using the facet table: The facet table organizes all the loaded datasets into a table to allow the user to search for datasets using metadata terms. **Rows** represent samples and **columns** represent assay types. **Cells** with numbers represent the number of datasets for a particular sample and assay combination. For example, in the cell with numbers 11/130, 11 represents the number of datasets currently loaded on the browser, while 130 represents the total number of datasets available.

- [A] To display the facet table, click the Tracks button near the top of the browser to display the Tracks menu. Click the teal box at the top of the menu to display the facet table.

- ➢ [B] To search for datasets, click on **Fetal Cells/Tissues** and then select **Fetal Brain**. Click the cell corresponding to the **Epigenetic Mark** column and the **Fetal Brain** row to list the datasets in this category.
- ➢ [C] To add the datasets, first select the datasets **H3K9me3 of Fetal Brain** and **H3K4me3 of Fetal Brain**. Then, click the **Add 2 tracks** button.
- ➢ To exit this window, click the **X** at the top-right of the panel or press **Esc**.



⊞ Epigenetic Mark	⊞ Expression	⊞ Long Range Interaction	⊞ Other Assays	⊞ Transcription Regulator	
0/1225	0/80	n/a	n/a	n/a	⊞ Adult Cells/Tissues
0/4	n/a	n/a	n/a	n/a	⊞ Cancer Cells
0/703	0/30	n/a	n/a	n/a	⊞ ES/iPS Cells
					⊟ Fetal Cells/Tissues
0/14	n/a	n/a	n/a	n/a	⊞ Fetal Adrenal Gland
					⊟ Fetal Brain
0/32	0/4	n/a	n/a	n/a	Fetal Brain
0/4	n/a	n/a	n/a	n/a	brain, fetal day122 U
0/1	n/a	n/a	n/a	n/a	brain, fetal day112 U
0/1	n/a	n/a	n/a	n/a	brain, fetal day117 F
0/3	n/a	n/a	n/a	n/a	brain, fetal day122 M
0/2	n/a	n/a	n/a	n/a	brain, fetal day85 F
0/3	n/a	n/a	n/a	n/a	brain, fetal day96 F
0/1	n/a	n/a	n/a	n/a	brain, fetal day142 F
0/3	n/a	n/a	n/a	n/a	brain, dorsal neocortex, fetal
0/30	n/a	n/a	n/a	n/a	⊞ Fetal Heart
0/27	n/a	n/a	n/a	n/a	⊞ Fetal Intestine, Large
0/23	n/a	n/a	n/a	n/a	⊞ Fetal Intestine, Small
0/73	n/a	n/a	n/a	n/a	⊞ Fetal Kidney
n/a	n/a	n/a	n/a	n/a	⊞ Fetal Liver
11/130	0/12	n/a	n/a	n/a	⊞ Fetal Lung
n/a	n/a	n/a	n/a	n/a	⊞ Fetal Membrane
0/72	0/31	n/a	n/a	n/a	⊞ Fetal Muscle
0/1	0/1	n/a	n/a	n/a	⊞ Fetal Ovary
0/14	n/a	n/a	n/a	n/a	⊞ Fetal Placenta
0/120	0/18	n/a	n/a	n/a	⊞ Fetal Skin
0/5	0/2	n/a	n/a	n/a	⊞ Fetal Spinal Cord
0/1	n/a	n/a	n/a	n/a	⊞ Fetal Spleen
0/22	n/a	n/a	n/a	n/a	⊞ Fetal Stomach
0/2	n/a	n/a	n/a	n/a	⊞ Fetal Testes
0/21	n/a	n/a	n/a	n/a	⊞ Fetal Thymus
0/14	0/8	n/a	n/a	n/a	⊞ Placenta

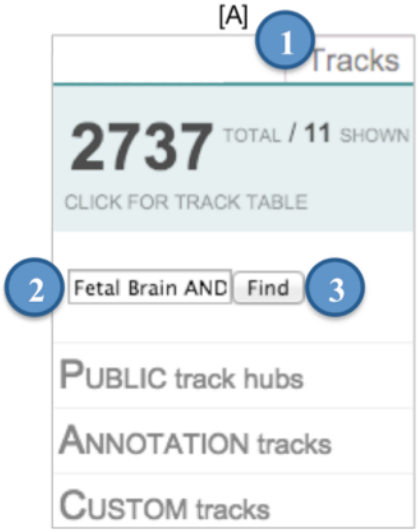
To get additional tracks, load public track hubs.

H3K9me3 of Fetal Brain	bigwig	i
DNase hypersensitivity of Feta...	bigwig	i
H3K4me3 of Fetal Brain	bigwig	i
H3K27me3 of Fetal Brain	bigwig	i
ChIP-Seq input of Fetal Brain	bigwig	i
DNase hypersensitivity of Feta...	bigwig	i
DNase hypersensitivity of Feta...	bigwig	i
DNase hypersensitivity of Feta...	bigwig	i

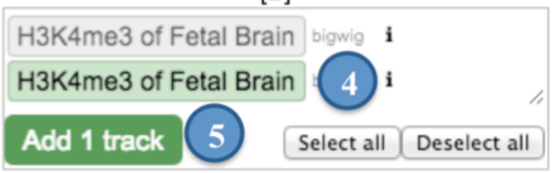
3.3.2. Searching for datasets using the track search box

- ➢ [A] Click the **Tracks** button. In the search box, type **Fetal Brain AND H3K4me3**, and then click **Find**. The **AND** is a logical operator so must be uppercased, but search terms are otherwise case-insensitive.
- ➢ [B] This will list all replicates for this sample and assay type. Only one of the two boxes listed can be selected since the other replicate has already been loaded as indicated by the light gray box. Click the second **H3K4me3 and Fetal Brain** and then click **Add 1 track**.
- ➢ Repeat this process to add one more replicate track for **Fetal Brain AND H3K9me3** and click outside the floating window to get back to the browser**.**

[A]



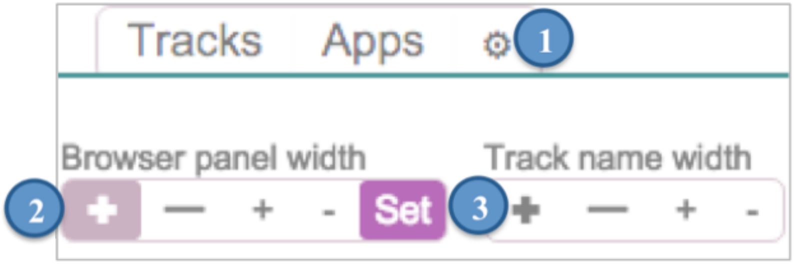
[B]



To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRworkshop in the Retrieve text box and click the status named Add datasets. For more details about sessions, refer to section 4.3 on page 13.

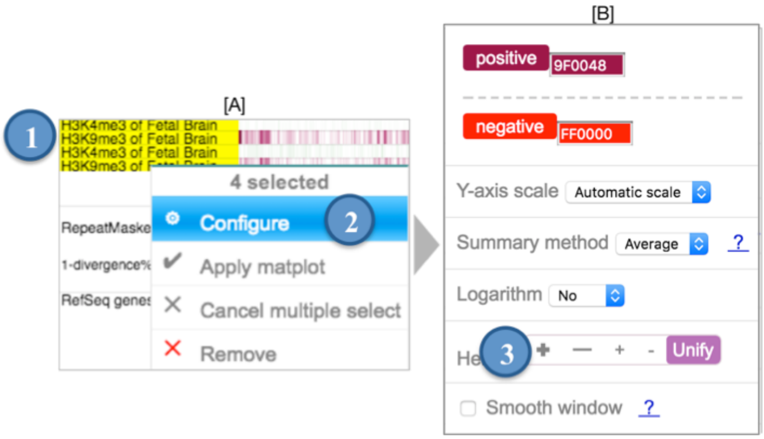
3.3.3. Formatting the view range

- ➢ To adjust the **browser width**, click the ⚙ button near the top of the browser. The + and - buttons increase or decrease the browser width, respectively. Click **+** then **Set** to change the browser width.
- ➢ Similarly, the **width of the track names** can be changed using the **+** and **-** buttons.

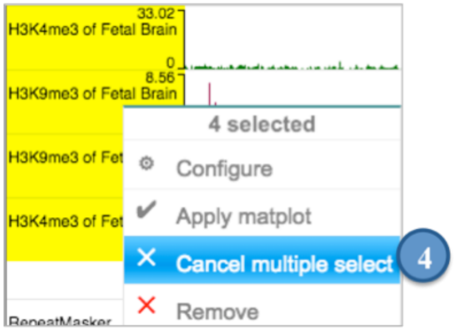


3.3.4. Changing the height of tracks

- ➢ [A] To select multiple tracks, hold down the **shift key** and click the names of the **four Fetal Brain tracks** that were just added. The track name will be highlighted in yellow when selected. Right-click the selected tracks then click **Configure**.
- ➢ [B] Click the **+** sign repeatedly to increase the height of the tracks. This does not change the y-axis scale.
- ➢ Click outside the configuration menu to close the menu.



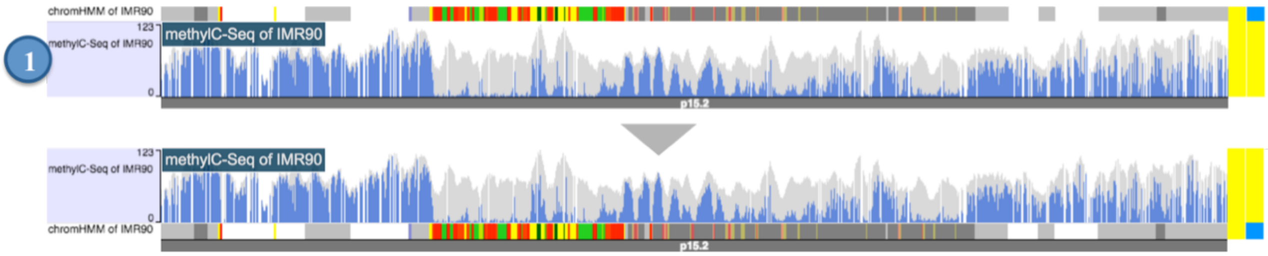
- ➢ To deselect the four tracks, right-click the highlighted track names and select **Cancel multiple select**.



3.3.5. Reordering tracks

3.3.5.1 Reordering tracks by clicking-and-dragging

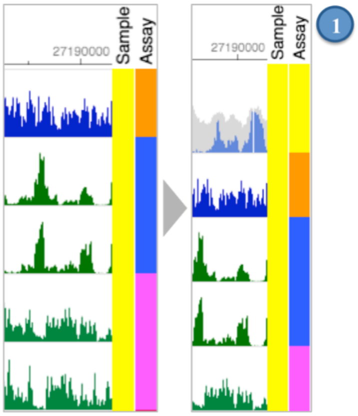
- ➢ Tracks can be reordered by clicking on a track name and moving that track to a new position.
- ➢ Select the **methylC-Seq of IMR90** track and move the track above the **chromHMM** track.



To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRworkshop in the Retrieve text box and click the status named Configure browser 1. For more details about sessions, refer to section 4.3 on page 13.

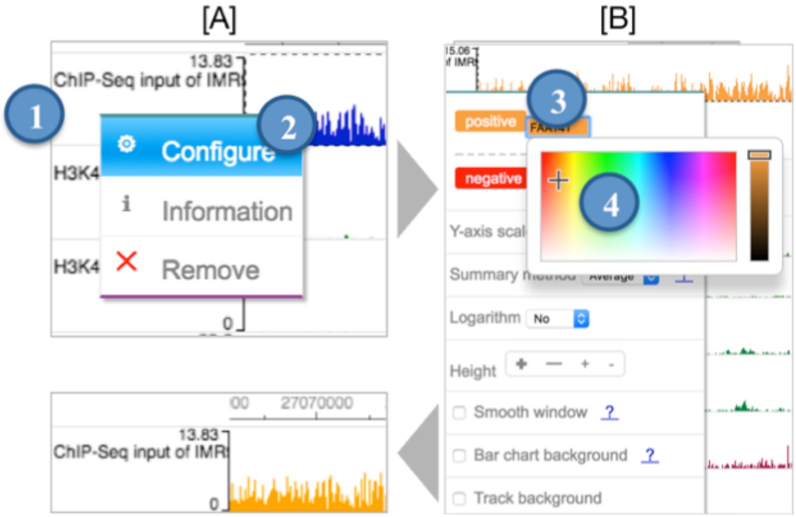
3.3.5.2 Reordering by metadata terms

- ➢ Alternatively, tracks can be reordered based on assay type, click the **Assay** term above the heatmap.



3.3.6. Changing the color of tracks

- ➢ [A] Select the **ChIP-Seq input of IMR90** track by right-clicking on the track name and then click **Configure**.
- ➢ [B] In the configuration menu, click the **positive** button and then select the **color** of your choice using the **color picker** to change the color of the track.



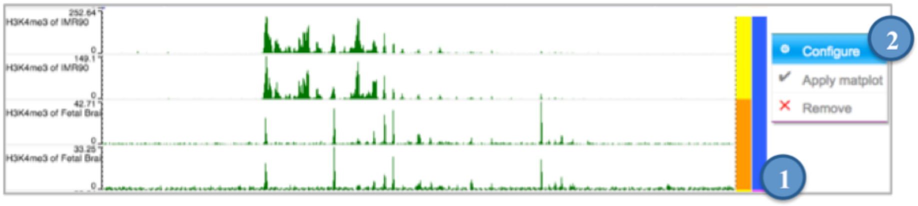
- ➢ Click outside the configuration menu to close the menu.

To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRworkshop in the Retrieve text box and click the status named Configure browser 2. For more details about sessions, refer to section 4.3 on page 13.

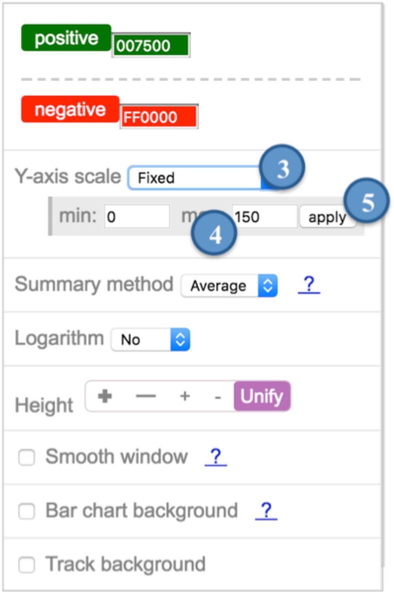
3.3.7. Changing the y-axis scale for tracks

- ➢ [A] To select the four H3K4me3 tracks, right-click the **blue rectangle** in the heatmap (under the Assay metadata term column) and click **Configure**.
- ➢ [B] To change the height of the selected tracks, click the **y-axis scale** drop-down menu and then select **Fixed**. This will generate textboxes in which a y-axis range can be entered. Type **150** in the max textbox. Click **apply** to change the y-axis scale.
- ➢ [C] To change the color used to mark regions where the data is beyond the range of the y-axis, click the **beyond threshold** button. This will generate a window with color options; select the **color** of your choice.

[A]



[B]


positive 007500

negative FF0000

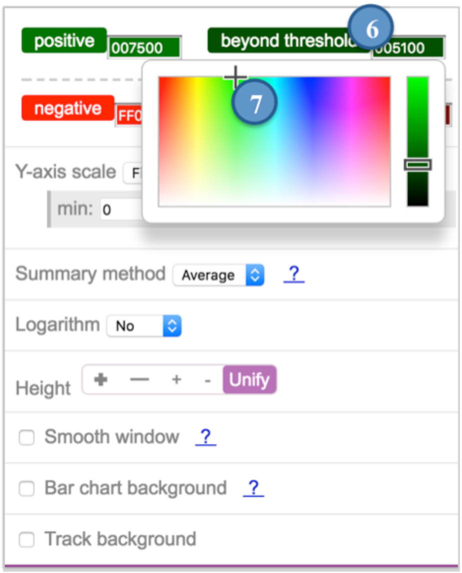
Y-axis scale Fixed

min: 0 max: 150 apply

Summary method Average

Logarithm No

[C]


positive 007500

beyond threshold 005100

negative FF0

Y-axis scale F

min: 0

Summary method Average

Logarithm No

- ➢ Click anywhere outside the **Configure** box to close the menu.

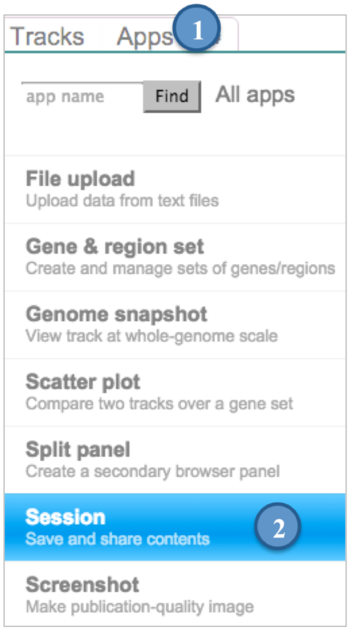
To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRworkshop in the Retrieve text box and click the status named Configure browser 3. For more details about sessions, refer to section 4.3 on page 13.

4. Saving, sharing, and retrieving EpiGenome Browser sessions

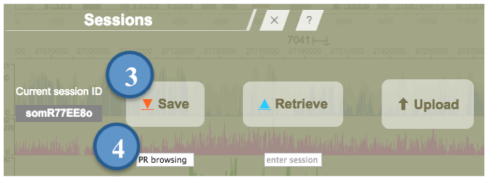
4.1. Saving sessions

- ➢ [A] Click the **Apps** button and then select **Session**.
- ➢ [B] Type **PR browsing** in the dialogue box under the Save button and then click **Save**. Click the **X** at the top-right of the panel to close the menu

[A]

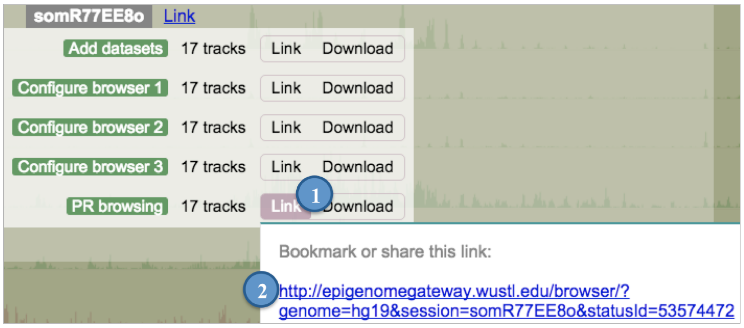


[B]



4.2. Sharing Links for Collaboration

- ➢ To generate the link for this session, click the **Link** button next to the session name.



Bookmark or share this link:

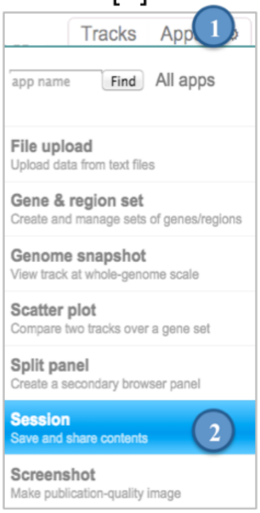
http://epigenomegateway.wustl.edu/browser/?genome=hg19&session=somR77EE8o&statusId=53574472

This will create a floating window with a **URL** that can be shared.

4.3. Retrieving Sessions

- Launch a new instance of the EpiGenome browser in a new window by going to http://epigenomegateway.wustl.edu/browser. Select the **Human hg19** genome assembly. Click the **GENOME Browser »** button to proceed to the browser.
- [A] Click the **Apps** button, and then select **Session**.
- [B] Enter the session ID PRdemo in the search box under the Retrieve button, and then click **Retrieve**. This will list all the available statuses under this session ID. Click on the Start status. This will retrieve a new session that will be used for the last section of this tutorial.

[A]



[B]

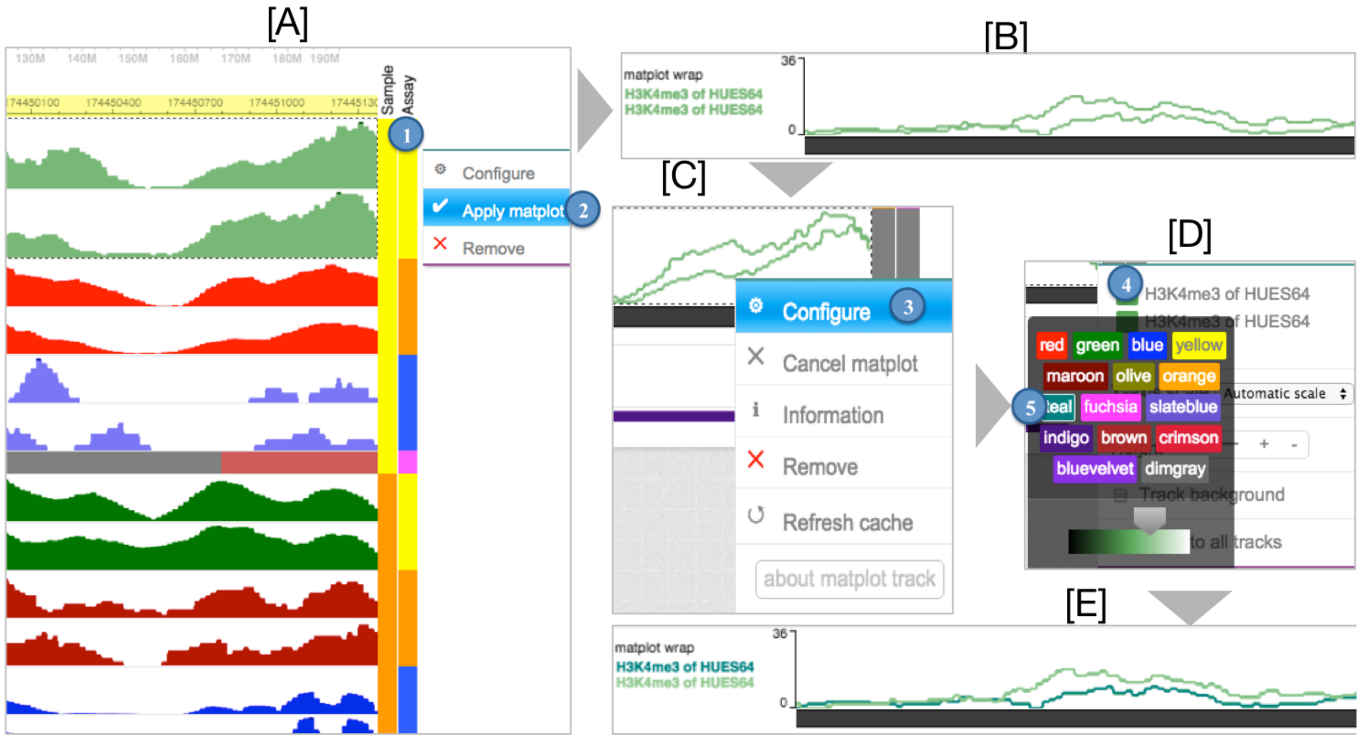
- Click the **X** at the top-right of the panel or press **Esc** to close the menu.
- In this session, we will be using Roadmap human HUES64 cells and left ventricle tissue data to analyze bivalent promoters of genes associated with heart development.
- The Start session contains pre-loaded and formatted H3K4me3, H3K27me3, RNA-Seq, and chromHMM tracks for the left ventricle (heart) and ESC line HUES64.

5. Using Apps on the EpiGenome Browse

5.1. Matplot

Matplot: A matplot (line plot) is used to compare two or more **numerical** tracks by plotting datasets on the same **X** and **Y axes**. Move to the gene ***HAND2*** in the browser.

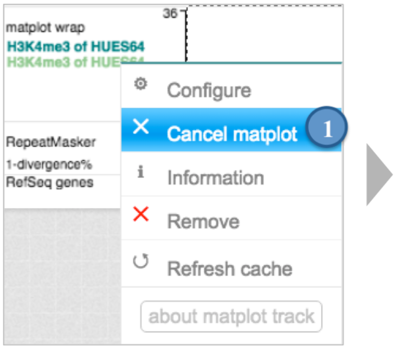
- ➢ [A] **Right-click** the first yellow box in the Assay column of the metadata heatmap (H3K4me3 of HUES64) and click **Apply matplot**.
- ➢ [B] This will generate a matplot track.
- ➢ [C] To change the color of the numerical tracks, **right-click** the name of the matplot track and click **Configure**.
- ➢ [D] Click the **green-box** beside the first track. Select **teal** to change the color of light green to teal.
- ➢ [E] This results in a track where the H3K4me3 data from the two HUES64 replicates share the same y-axis and can be easily compared. It is evident from this view that these tracks both show the same pattern of H3K4me3 across the *HAND2* gene.



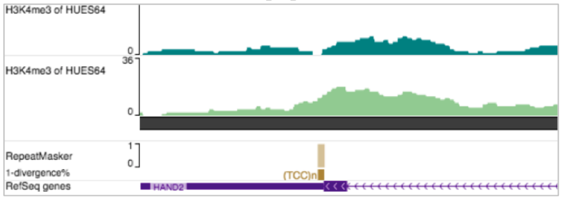
To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named Matplot. For more details about sessions, refer to section 4.3 on page 13.

- [A] To exit from the matplot view, **right-click** the matplot track name and click **Cancel matplot**.
- [B] This ungroups the tracks, but keeps their new colors. These tracks can then be moved back to the original order.

[A]



[B]



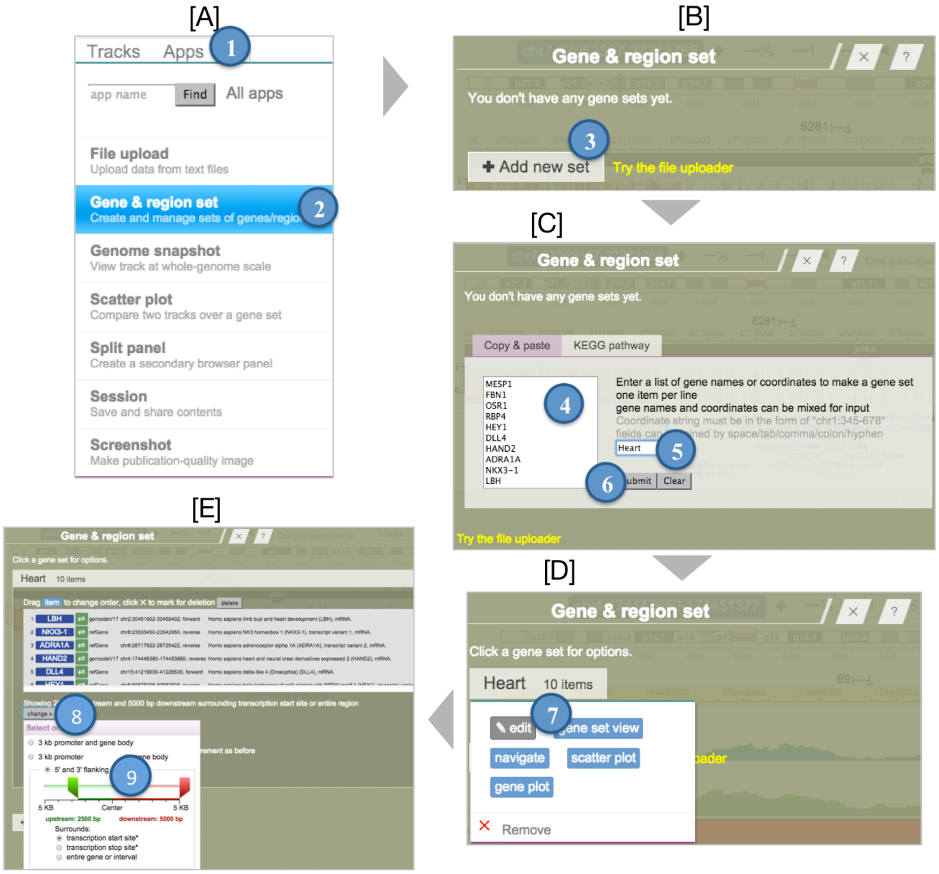
5.2. **Gene & region set:** The Gene & region set app is used to visualize multiple genomic regions in parallel. The app enables track data to be displayed over nonadjacent regions on a chromosome or even regions on different chromosomes. We are going to use this app to look at a couple of genes with bivalent promoters in the ESC line HUES64, but active transcription in the left ventricle sample.

- Copy the 10 genes in the table, these will serve as the Heart Gene Set.

MESP1
FBN1
OSR1
RBP4
HEY1
DLL4
HAND2
ADRA1A
NKX3-1
LBH

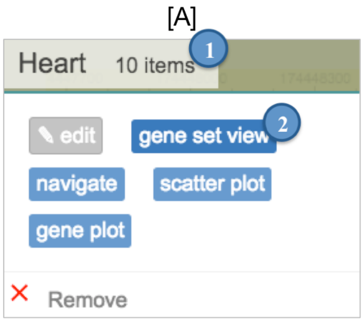
- [A] Click the **Apps** button and then select **Gene & region set**.
- [B] To add a set of genes click **+ Add new set**.
- [C] Paste the 10 genes into the large textbox, name the set “Heart”, and click **submit**.
- [D] Select the **Heart** button and click **✎ edit**.
- [E] Under the list of genes, click the **change »** button to modify the view range. First select the radio button for **5’ and 3’ flanking** and then select the radio button for

transcription start site*. Adjust the **green** slider to 2.5 KB and the **red** to 5 KB. Click outside the window to exit this window.



To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named Gene Set. For more details about sessions, refer to section 4.3 on page 13.

- [A] To view the “Heart” gene set go to apps **Gene & region set**, select the **Heart** button, and then click **gene set view**.



- Change the y-axis scale on all RNA-Seq tracks to a fixed scale with a max of 150.
- Reconfigure the track name width to see the full name of each track.
- This results in a tiled view of 10 genes arranged beside each other. By setting the same y-axis scale for all RNA-seq data, the expression of each gene in the set can be better appreciated. Additionally, looking closely at each gene reveals that the HUES64 tracks

have strong H3K4me3 peaks over the TSS and promoter region along with H3K27me3, but the H3K27me3 signal is not seen in the left ventricle tracks.

To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named Gene Set View. For more details about sessions, refer to section 4.3 on page 13.

- To exit the gene set view, click the **Showing entire set | X** button near the top left side of the browser.

5.3. Navigate regions: The navigate regions app allows the user to navigate across the regions specified in a gene & region set.

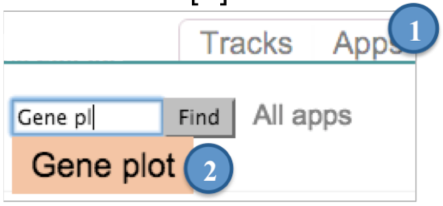
- [A] Click the **Apps** button and then select **Gene & region set**.
- [B] Select the **Heart** button and click **navigate**.
- [C] A panel will now be displayed with the 10 genes in the Heart gene set. Click **HAND2** to navigate to the *HAND2* gene.

To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named Navigate. For more details about sessions, refer to section 4.3 on page 13.

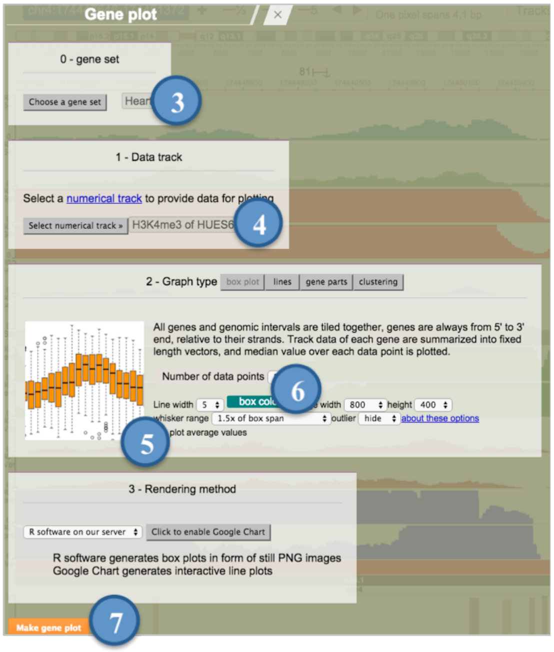
5.4. **Gene plot:** The Gene plot app summarizes the data distribution across multiple regions by splitting the regions into bins.

- Change the gene view area in each Gene & region set to **gene body only**.
- [A] Click the **Apps** button. In the search box, type **Gene plot** and select the **Gene plot** app. This app can also be accessed by clicking on a gene set.
- [B] To make a gene plot, follow these steps:
  - In the **0 – Gene set** section, click the **Choose a gene set** button. Select the **Heart** gene set.
  - In the **1 – Data track** section, click the **Select numerical track »** button and then click **H3K4me3 of HUES64**.
  - In the 2 **– Graph type** section, use the default graph type, a box plot. Click the **plot average values** check-box. To change the box color, click the **box color** button and then select the color of your choice.
  - Click the **Make gene plot** button to create the gene plot.

[A]



[B]



Select a numerical track to provide data for plotting

All genes and genomic intervals are tiled together, genes are always from 5' to 3' end, relative to their strands. Track data of each gene are summarized into fixed length vectors, and median value over each data point is plotted.

Number of data points

Line width 5 box colo width 800 height 400

whisker range 1.5x of box span outlier hide about these options

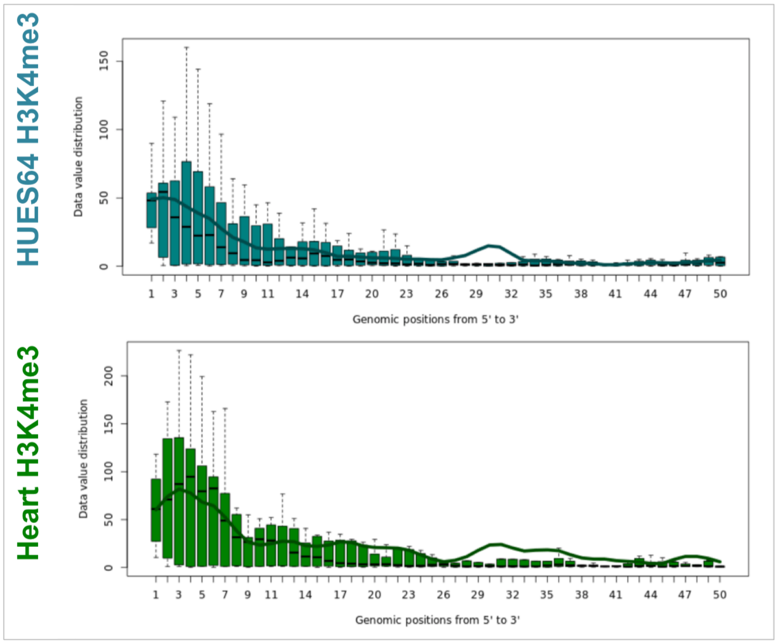
plot average values

R software on our server

R software generates box plots in form of still PNG images
Google Chart generates interactive line plots

- Repeat these steps with the Heart gene set and the **H3K4me3 of Left Ventricle** dataset. Use a different color for the boxplot.

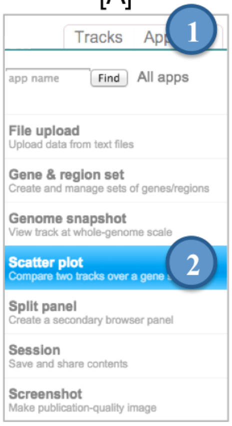
- Comparison of the two gene plots reveals that the H3K4me3 signal is the strongest in the 5’ upstream region of the gene with an appreciable depletion when moving towards the 3’ end of the gene.



5.5. **Scatter plot:** The Scatter plot app is used to assess the relationship between two different datasets across multiple genomic regions.

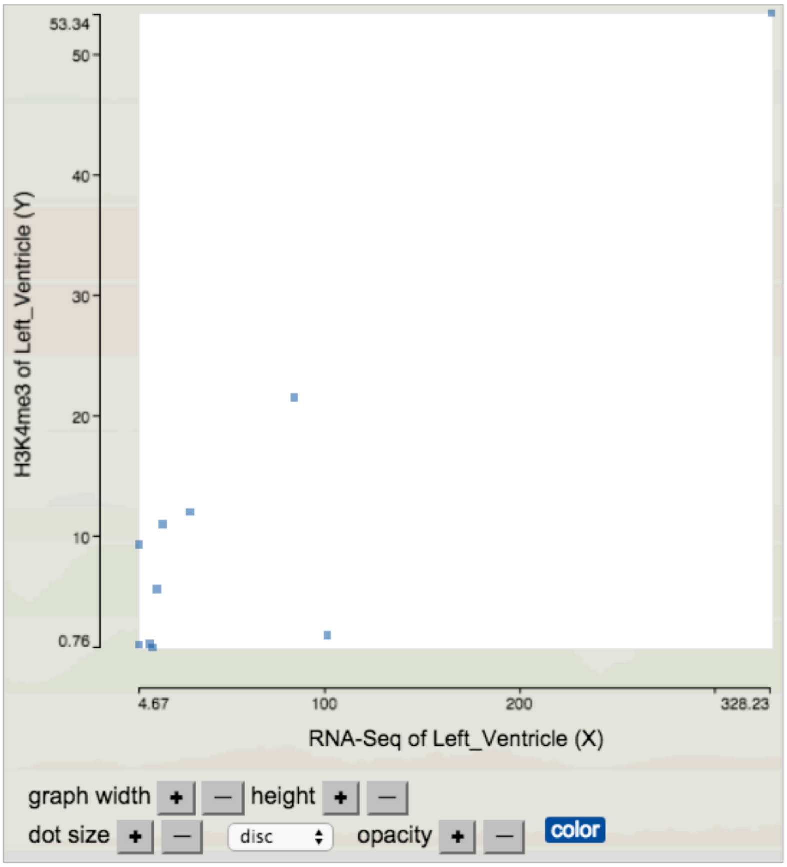
- [A] Click the **Apps** button and then select **Scatter plot**.
- [B] To select the genomic regions to plot, click the **Choose a gene set** button and then click the gene set **Heart**. To select the numerical track to plot on the x-axis, first click the **Choose track »** button above the **for X axis** button then select the second **RNA-Seq of Left_Ventricle** track. To select the numerical track to plot on the y-axis, first click the **Choose track »** button above the **for Y axis button** then select the second **H3K4me3 of Left_Ventricle** track. Click **SUBMIT** to create the plot.

[A]



[B]

- The resulting scatter plot shows the correlation between gene expression and H3K4me3 for the left ventricle across the gene set. H3K4me3 is associated with active promoters, so the correlation is expected.



- To exit the app, click the **X** at the top-right of the panel or press **Esc**.

6. Creating Custom Tracks on the EpiGenome Browser

Now that you are more familiar with the WashU EpiGenome Browser it is time to explore your data.

6.1. Creating custom tracks

- [A] Click the **Tracks** button then click **Custom Tracks**.
- [B] Click the **+ Add new tracks** button.
- [C] The browser supports many file formats for custom tracks. Data must be hosted on a web server accessible to the browser. Click **BED** to create your first custom track.
- [D] Click **Use example** and then **SUBMIT**. The track now appears in the browser window, click **X** to return to the browser window.

Tracks need to be hosted on a web server that is accessible by this browser server.

Got text files instead? Upload them from your computer.

To submit methylC or matplot track, use Datahub.

BED file URL: http://vizhub.wustl.edu/hubSample/hg1

Track name: example bed track

Show as: full

To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named Custom Track. For more details about sessions, refer to section 4.3 on page 13.

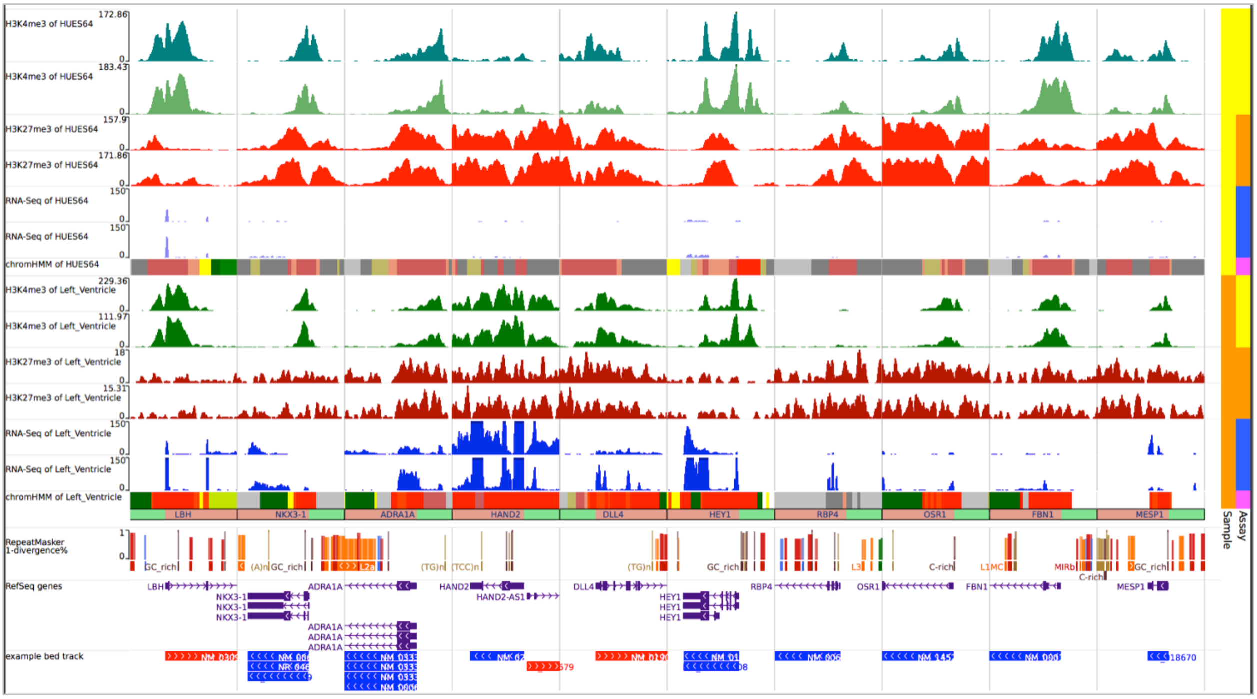
6.2. Saving EpiGenome Browser screenshots

- [A] Finally, you may want to take a screenshot of your final tracks, a gene of interest, or a gene set. Reapply the Heart gene set view with 2.5 KB upstream and 5 KB downstream of the TSS. Click the **Apps** button then click **Screenshot**.
- [B] Click the **show track name** checkbox then click **Take screenshot**. Click either the **Link to the svg file** button or the **Link to the pdf file** button to save a high quality screenshot.

[A]

[B]

This generates an SVG file that can be printed to PDF format with your web browser



To catch up to this point of the workshop, click on the Apps menu and then select Session. Enter the session ID PRdemo in the Retrieve text box and click the status named End. For more details about sessions, refer to section 4.3 on page 13.

7. Training and support

- For more documentation and tutorials, visit: http://epigenomegateway.wustl.edu/support/index.html
- Follow the WashU EpiGenome Browser on

http://epigenomegateway.wustl.edu/+

http://epigenomegateway.wustl.edu/fb

https://twitter.com/wuepgg

https://github.com/epgg/eg