Tutorial: Jump Start on the Human Epigenome Browser at Washington University

This brief tutorial aims to introduce some of the basic features of the Human Epigenome Browser, allowing users to navigate the browser and interact with data tracks in a meaningful way. In order to access the public site, please visit http://VizHub.wustl.edu/. In this tutorial, "heatmap" will be used to refer to "genome heatmap" (panel A in **Supplementary Figure 2**), and "control panel" will be used to refer to panel G in **Supplementary Figure 2**.

Help desk

Help desk for the Epigenome browser is at
http://epigenome.wustl.edu/subtleKnife/helpdesk.html

Please check the site for most updated tutorial and video demo.

Fine-tuning genome heatmap display

The color of individual heatmap cells is determined by the numerical data associated with a cell and a threshold value for that data track. To understand how the threshold value is computed, go to the tool box which floats on the web page (it can be moved by dragging the banner on the top):
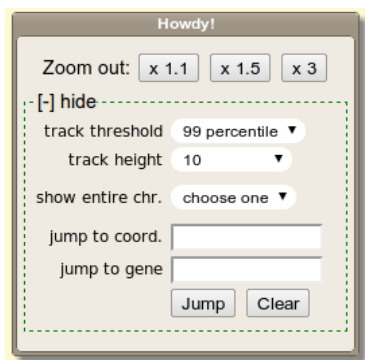

The drop-down menu for the "track threshold" gives options for computing the track threshold score. If "max" is selected, the maximum value for each track at the currently displayed region will be used as the threshold value. Otherwise, the user can select a percentile value to be the threshold.

Once a method is selected to compute track thresholds, the heatmap cell color will be decided accordingly. This can be visualized using the context menu accessed by right-clicking on a single data track in the heatmap:
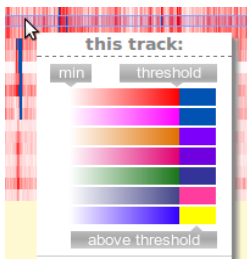
Each row of colored bars represents a different color scheme that can be assigned to a data track by the user. If the cell value is below the threshold score, the color of the cell will be determined according to the color gradient leading up to the threshold. As a value approaches the threshold, the color assigned to the cell will be progressively darker. If a value is above the threshold by any amount, the cell will be assigned a distinct color.

The height of each row in the heatmap can be changed using the "track height" option in the toolbox. If track height is greater than 20 pixels, the heatmap display will be replaced by a "wiggle plot", where vertical lines indicate the relative heatmap cell value:
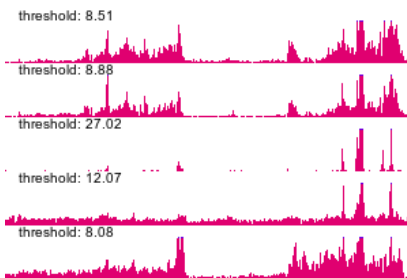
Finally, tracks in the heatmap can be rearranged in various ways to make visualization more intuitive for the individual user. Click and drag the striped bar corresponding to a data track, found to the right of metadata heatmap, in order to move a single track. Group of tracks can be moved together by dragging the colored blocks in metadata heatmap. Tracks can be sorted according to metadata terms by clicking on the column labels in the metadata heatmap. When the correlation function is turned on, tracks can be sorted according to correlation coeffients by clicking on the coefficients image to the left of the heatmap.

Navigation

To view a wider genomic region in the heatmap, click the "zoom out" button in the toolbox; to "zoom in", click on the chromosome ideogram below the heatmap and drag the cursor to the desired genomic coordinate. The heatmap will zoom to display the selected region down to single nucleotide views. To move left of right within the current zoom level, click, hold and drag the heatmap in either direction. Please wait for the loading image to disappear before dragging the heatmap further. Depending on the speed of internet connection and the number of tracks displayed, this waiting time can be somewhere between several milliseconds to two seconds.

Customizing metadata color-coding area

Columns in the metadata heatmap correspond to the metadata terms. In order to add/remove terms or browse the collection of terms, go to the "Metadata" tab in the control panel. Terms

with "/" icon can be expanded or collapsed by clicking to show or hide children terms. Use the check boxes to add or remove columns from the metadata heatmap:
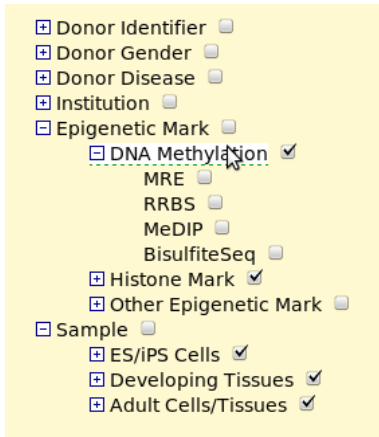

As an example, the metadata terms chosen above will produce the following metadata heatmap:
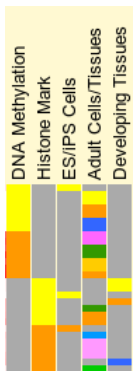

Clicking on term names above the heatmap will sort tracks according to the annotation of that attribute. Drag the term name to re-arrange metadata columns.

Track selection

Open the "Track management" tab in the control panel to access the track selection grid (**Supplementary Figure 2**). By default, rows of the grid correspond to sequencing experiment sample names, and columns correspond to the type of assay (epigenetic marks). Name with "/" icon means the marks can be clicked on to show/hide detailed terms. Numbers in the grid occur in pairs, indicating the number of tracks annotated by the combination of row/column terms, and number of tracks in this group currently on display. Click a pair of numbers, and a panel will appear in toolbox allowing the user to add/remove tracks with check boxes:
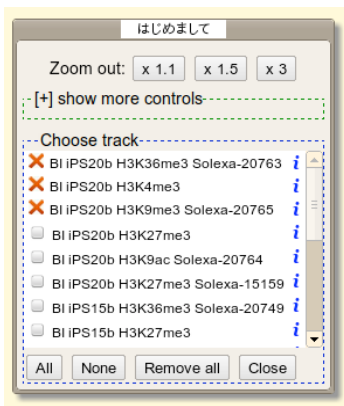
In this "Choose track" panel, tracks currently on display will be labeled with a red check. Click on it and the track will be removed from display. Click on "Remove all" to remove all tracks. Tracks that are available for but not on display will have a check box on the left of the track name. Check the box will result in the following panel:
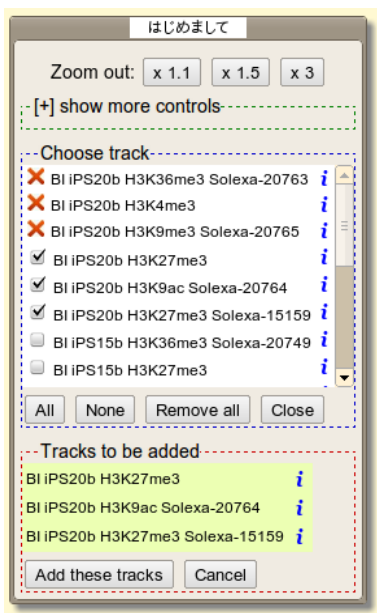

In this "Track to be added" panel, click on "Add these tracks" button to add the tracks for browsing. The user can make selections on multiple number pairs and add tracks all together.

To view detailed information for each track, click on the blue letter "i" following the track name and an additional panel with detailed track information will appear in toolbox. By clicking the "GEO" icon, the user will be taken to the record to this track in the GEO database:
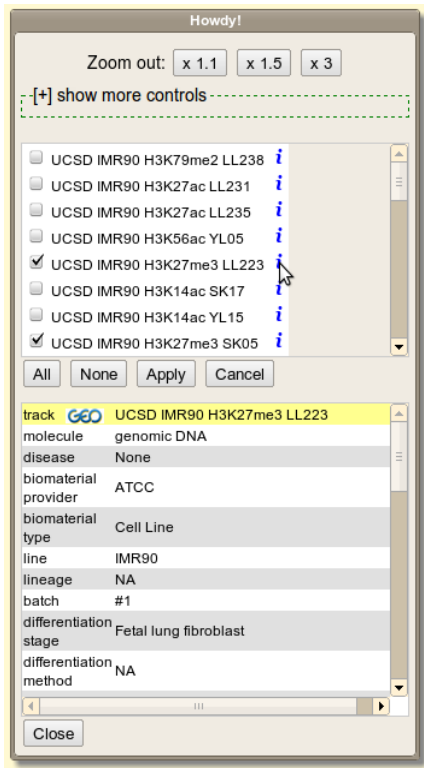
Short cuts are available to quickly select/remove all tracks annotated by one term. Scroll mouse over a term in the track selection box and a small menu will appear with options "Select all" and "Un-select all". Click on either option to take effect:
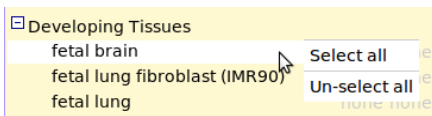
The track display can be cleared in a single click, using buttons at the bottom of the track selection grid.

Using the "Advanced options" tab on top of the track selection grid, the grid layout can be customized. The track selection panel can be reorganized using a single criteria. In this case, an indented text tree will be generated, and the user can click on the pair of numbers to select all tracks annotated by that term:

[-] hide

Define criteria to change layout of the track selection grid:

Criterion 1   Criterion 2 (optional)
[Sample   ▼]   [don't use   ▼]   [Generate track selection layout]

⊟ ES/iPS Cells **169**/**3**
    ⊞ Embryonic Stem Cells **131**/**2**
    ⊞ Induced Pluripotent Stem Cells **38**/**1**
⊟ Developing Tissues **113**/**4**
    fetal brain **11**/**1**
    fetal lung fibroblast (IMR90) **70**/**2**
    fetal lung **17**/0
    fetal kidney **8**/**1**
    fetal heart **6**/0
    fetal adrenal gland **1**/0
⊟ Adult Cells/Tissues **311**/**21**
    ⊞ Muscle **60**/**4**
    ⊞ Breast **50**/**7**
    ⊞ Kidney **5**/0
    ⊞ Liver **14**/0
    ⊞ Bone Marrow **6**/0
    ⊞ Fat **12**/0
    ⊞ Intestine **30**/0
    ⊞ Pancreatic Islet **8**/0
    ⊞ Blood **126**/**10**

Genomic feature tracks

Many types of genomic feature annotation are available under the "Genomic feature" tab in the control panel. Three plotting modes can be selected from (thin, full, density), and the gene structure will be plotted if the "gene body" track is turned to "full" mode. Once a genomic feature track is displayed, the user can right click on the track image to display a context menu for additional visualization options.

An example using genomic feature track display is shown below. Gene body is in full mode, and various repetitive elements are in thin mode. The region surrounding the *NSUN6* gene is shown. While SINE, LTR, LINE, DNA transposons are relatively uniformly distributed in this region, satellite DNA only exists inside the 8th intron of *NSUN6*, and that region has strong methylation signal (MeDIP experiment tracks).
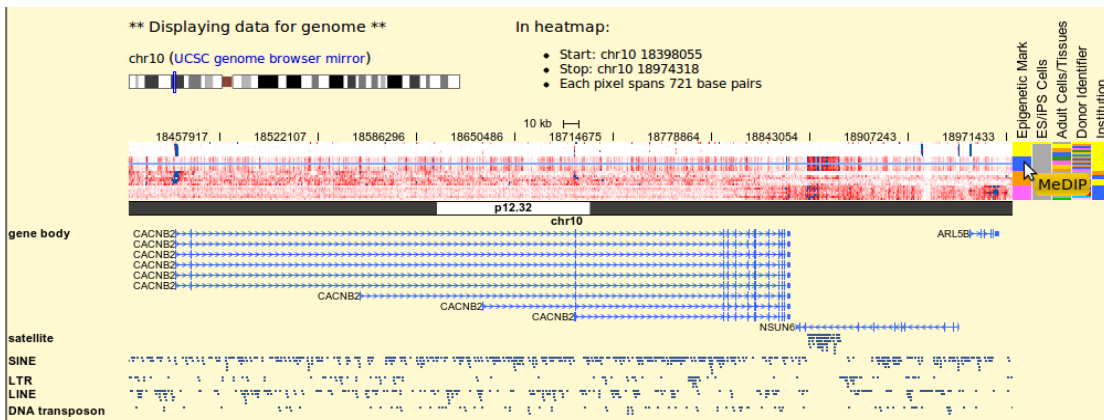
Correlation analysis

Open the "Correlation" tab in the control panel for correlation analysis options. Track scores can be correlated with genomic feature density data at the currently displayed region and an inter-track correlation can also be computed. Similar operations can also be applied to custom bigBed tracks. To activate these options, check the corresponding button in the control panel. Click the "Turn off correlation function" button to cancel correlation analysis.

Hypothesis test

Open the "Hypothesis test" tab in the control panel for options on this function. At least two groups of tracks must be defined to run a hypothesis test. In the panel shown below, "GROUP 1" and "GROUP 2" are two group holders provided by default, which can be modified or removed by the user.

To assign a new track to a group, the user can use the drop-down menus in the box on the right side of screen shot. Entries denoting tracks will appear in group boxes. To remove a track from selection, click the red cross image in the group box. A track can be assigned to a group using the context menu by right-clicking on the heatmap. Likewise, a group of tracks can be assigned to a group by right-clicking on the metadata heatmap. Click the "clear" button to remove all track assignments:
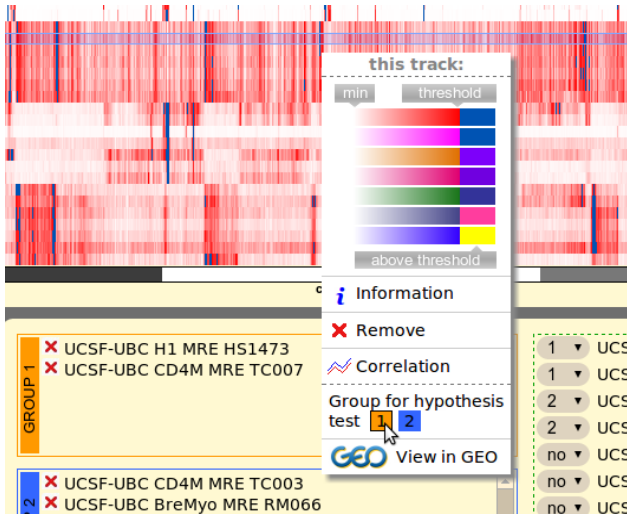
Click the "Run hypothesis test" button and the server will run the test. Once done, vertical bars corresponding to P-values will be plotted in a track below the chromosome ideogram. P-values less than or equal to the significance cutoff are plotted in red:
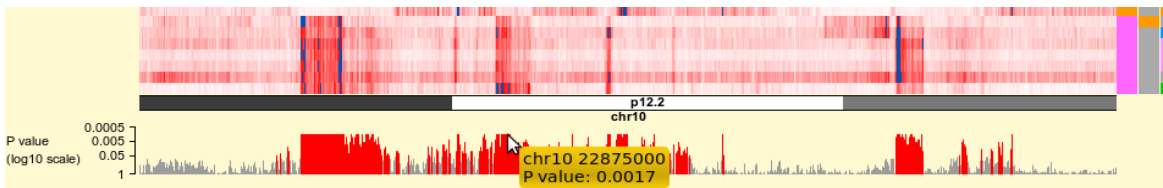
Multiple test correction can be applied to correct P-values. Select a correction option using the drop-down menu and click the "Run hypothesis test" button to run test again to generate corrected P-values.

Custom bigWig track

Custom bigWig tracks can be used to display numerical data across the genome. The bigWig format file can be generated using the "wigToBigWig" and "bedGraphToBigWig" utility programs available in the Kent Source Tree. The user needs to set up a HTTP web server and place this file on the web server directory.

To submit the custom bigWig track, open the "Custom track" tab in the control panel, click on "Submit bigWig" button, enter the bigWig file URL and a track name, and press "Submit" button to submit without specifying metadata for this track:

Submit a bigWig track

Enter bigWig file URL `http://mywebserver/sample.bigWig`

Track name `my wiggle track`

(Optional) annotate bigWig track using metadata:

[+] use our metadata vocabulary

[+] create/use your own metadata

Submit Cancel

The track will be displayed in the heatmap in the default color, blue. To change the color of the track, right click on the track to display the context menu and select a color scheme by clicking a row of colored bars. The track will also be assigned a row in the metadata heatmap. If no metadata for the track is provided, the cell will be colored gray.

To annotate custom bigWig tracks with metadata, the user can use the metadata vocabulary provided by the browser. Open this "Metadata" box to add annotation terms to the track using the drop-down menu:

[-] hide

- for metadata terms relevant to your custom track, select a value from the drop-down menu,
- don't make selection for irrelevant terms,
- for terms with additional detail, click ⊞ / ⊟ to show/hide details.

|-----Donor Identifier `select one`
|-----Donor Gender `select one`
|-----Institution `select one`
|-----Donor Ethnicity `select one`
|- ⊞ Epigenetic Mark
|- ⊟ Sample
　　　|- ⊟ ES/iPS Cells
　　　　　　|- ⊞ Embryonic Stem Cells `H1`
　　　　　　|-----Induced Pluripotent Stem Cells `select one`
　　　　　　|-----Derived Cells `select one`
　　　|-----Developing Tissues `select one`
　　　|- ⊞ Adult Cells/Tissues `select one`

If suitable terms are not available in the existing vocabulary, the user can create his/her own metadata term with a number of attributes, and annotate custom bigWig track using it. Open the corresponding box, click "Add new term" and a new panel will be displayed:
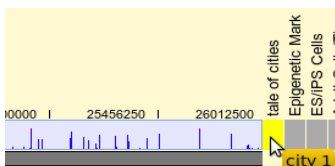
Enter the term name and attributes for this term to create a custom metadata term. Once added, the term will show up in a table ready to be used:

Following submission, the custom metadata term as well as the custom bigWig track will be displayed like the following:

Custom bigBed track

Users can display their collection of genomic features using a custom bigBed track. Genomic locations must be specified for the features, and the name/strand can be optionally specified. A special requirement is "id" on the 5th field of the bed file, in order for it to be compatible the drag-move feature of the Human Epigenome Browser. Open the bigBed custom track submission panel and enter the file URL and track name to submit.

Above is an example showing the appearance of a custom bigBed track with name "featureless genebody". Submitted custom tracks will occupy entries in the table seen at the bottom of screen shot, and clicking the red cross image will remove that track.

Data collation

Data collation is a unique feature of the Human Epigenome Browser, allowing the user to re-organize the heatmap view to concentrate on specific genomic features. To collate track data on a specific genomic feature, open the "Data collation" tab in the control panel. Under the "functional genomic features" tab, click the button next to the genomic feature to be used for collation. If this genomic feature is already on display, check the "Collate" option in the context menu by right-clicking on the track:
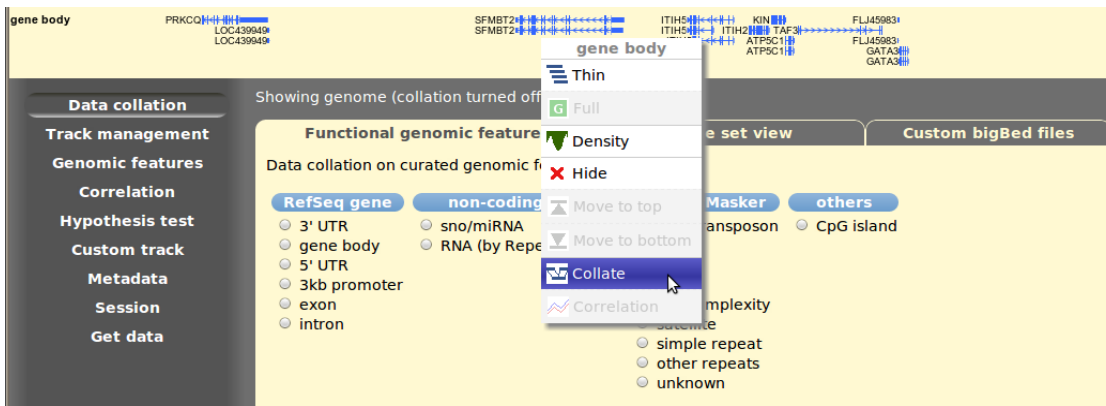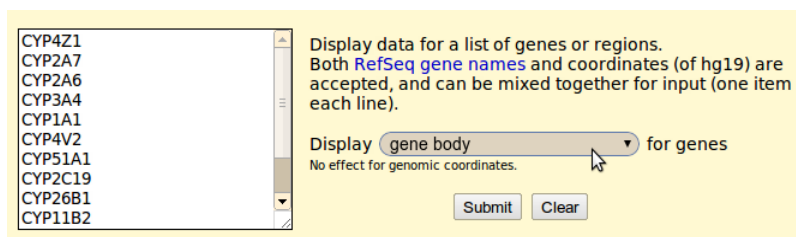

Following collation, the browser view will be similar to the following:

In the above example, data collation is performed on genes, and all the intergenic regions are removed as seen in the "gene body" track. A gray line is drawn between adjacent and non-overlapping genes, and will not be drawn between overlapping genes. All navigation operations still apply to this view (move/zoom). Click the "cancel collation" button to return to the genome view.

Gene set view

Data collation on genomic features will use features from part of the genome to organize the view. The features' original orientation and relative positions are all preserved. Conversely, using the gene set view, a set of genes can be used to organize the view irrespective of their original positions in genome. Open the "Gene set view" tab, click the "Custom gene set" button, and the following form will appear:


A list of gene names or genomic coordinates can be used as input. By default, a short list of cytochrome P450 genes is provided in no particular order. Following submission, the heatmap view will look like the following:
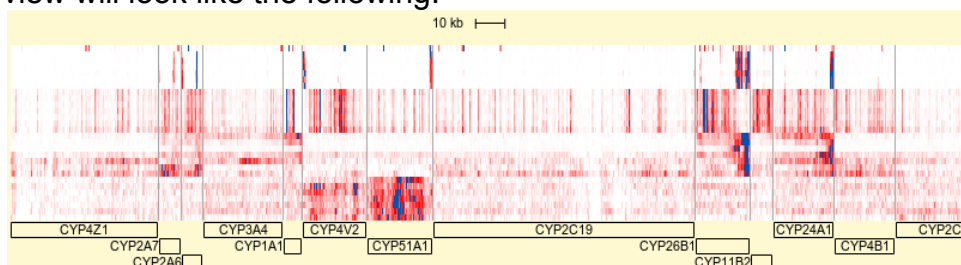

In this view, the chromosome ideogram below the heatmap is replaced by boxes. Each box represents one item (gene or genomic coordinate) submitted in the gene set, with the name inside of or adjacent to the box. The width of each box represents the relative length of the item. The "Zoom in" operation can be performed on this image by clicking on the image and dragging the cursor over the region of interest.

When submitting gene sets, the user can define which gene features parts should be displayed using the drop-down menu. Ready-made gene features can be selected for viewing, including 3 kb promoter (with or without gene body) and 5' or 3' UTRs, of the user can define a custom region. This allows the user to select a generic region around relevant gene features (i.e. transcription start/stop sites, or gene body) by defining the upstream and downstream sequence lengths. Upon choosing this option, a small control panel will appear. Drag the colored gliders to define the length of upstream/downstream regions to display:
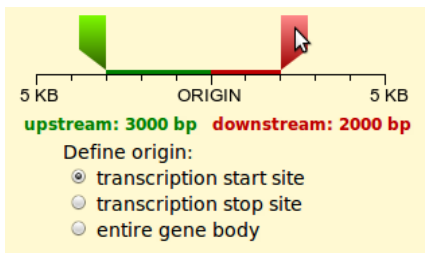
Following gene set submission, the control panel will contain the list of submitted genes and a brief detail for review. Also shown is a gene set editing panel. The user can change which gene features to display of sort the gene set.   Clicking the "remove all genes" button will clear this gene list and return to the genome view:
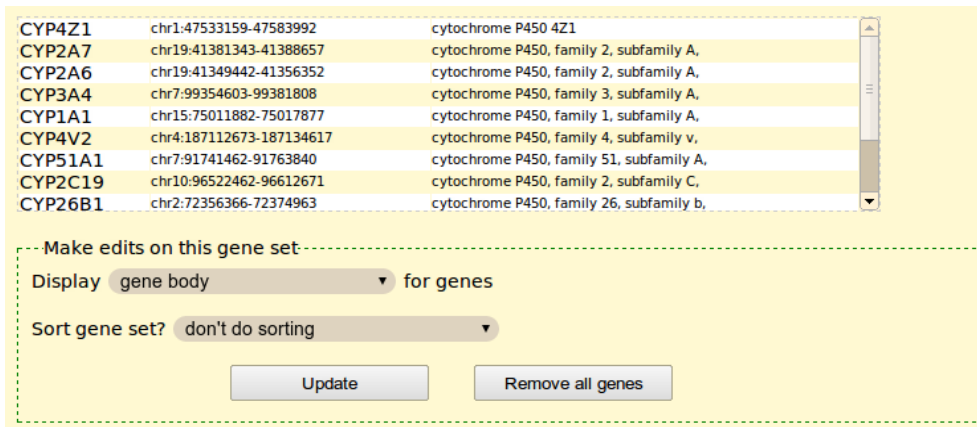

The gene set can be sorted according to the mean track signal. By selecting the "by average track score" option in the "sort gene set" drop down menu, an additional drop-down menu of track names will appear. The user can then select a track and click the "Update" button to sort by the average score of the selected track:
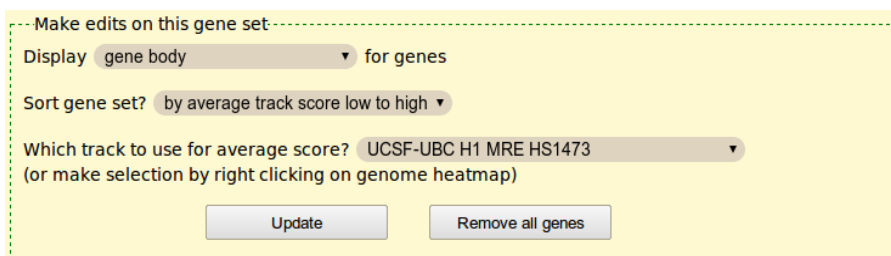
More conveniently, the user can right click on the heatmap to select a track for sorting (after selection, the user must click the "Update" button to sort):

13
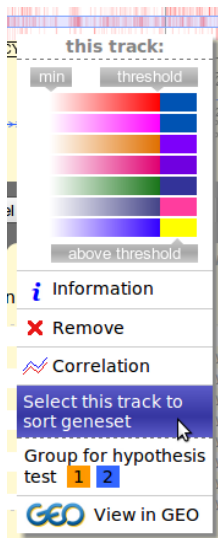
The following is the view after sorting the gene set by the average track score of the last track (in green):
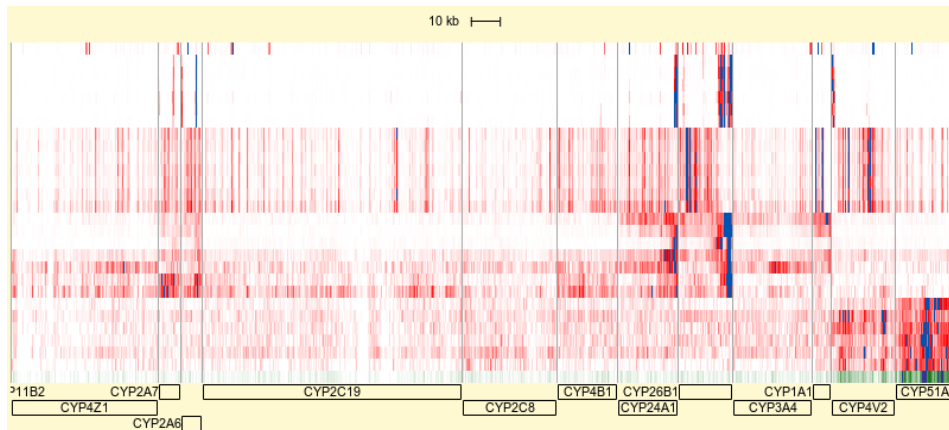
KEGG pathway view

The Kyoto Encyclopedia of Genes and Genomes (KEGG, http://www.genome.jp/kegg/) database contains carefully curated information on metabolic pathways, enzyme-coding genes, and metabolites in all genome-sequenced species. With the Human Epigenome Browser users can view track data on genes in KEGG pathways with a few simple steps. Go to the "Data collation" panel, and click "KEGG pathway" button. The following panel will appear:

The user can directly enter KEGG pathway identifiers (e.g. path:hsa00010) and view data; however, the user will usually need to do keyword search for pathways of interest. By searching for the keyword "glyco", the following pathways will be found:

Click the pathway icon to view pathway details in the KEGG database. Click on the pathway link to perform data collation. The following is the view produced by data collation on the 65 genes in the glycolysis pathway (half of entire gene set is shown):
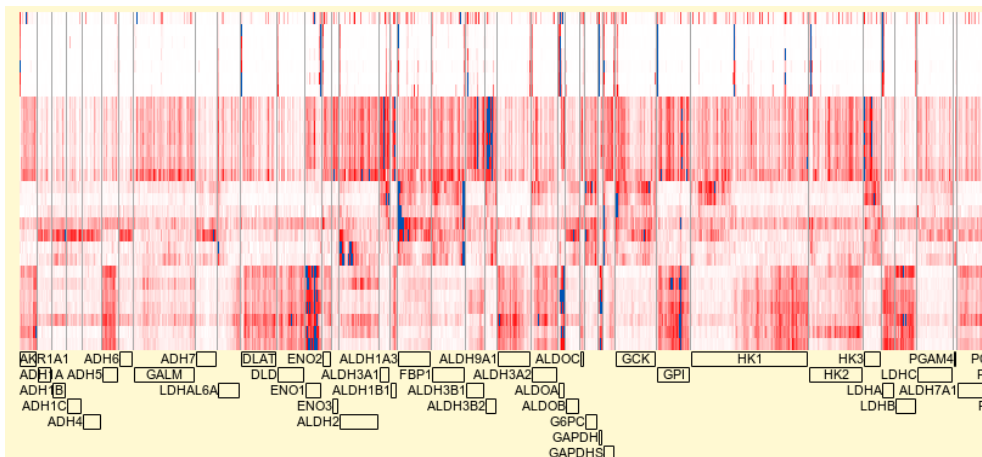
Session

Users can save their entire browsing status, including their custom tracks and metadata entries using the session function. Go to the "Session" tab, click the "Save status" button and the current browsing status will be saved under current session. Multiple statuses can be saved under one session, and an optional short text is available to describe each status:

Save the session ID and use it to retrieve saved statuses later or for sharing with colleagues. Enter the session ID into the second text field in the panel and click the "Retrieve status" button. All saved statuses will be displayed in the table. Previous browsing statuses can be restored by clicking on the status link.