HiGlass: Synchronized Exploration and Comparison of Multiple Genomic Datasets

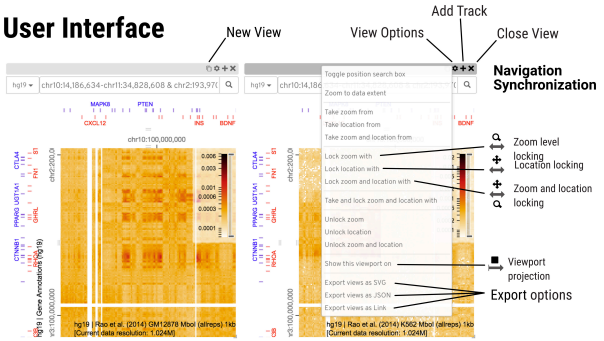
DEMO
http://higlass.io

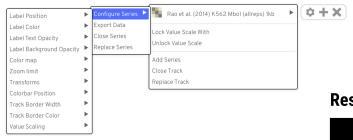
Peter Kerpedjiev[1], Nezar Abdennur[2], **Fritz Lekschas**[3], Chuck McCallum[1], Kasper Dinkla[3], Hendrik Strobelt[3], Jacob M Luber[1], Scott Oullette[1], Alaleh Azhir[1], Nikhil Kumar[1], Jeewon Hwang[1], Soohyun Lee[1], Burak H Alver[1], Hanspeter Pfister[3], Leonid A Mirny[2], Peter J Park[1], and Nils Gehlenborg[1]

[1]Harvard Medical School, [2]Massachusetts Institute of Technology, and [3]Harvard John A. Paulson School of Engineering and Applied Sciences

Abstract. HiGlass is an online viewer for Hi-C and other genomic data. It provides a customizable interface for arranging and synchronizing the navigation of multiple views across samples, loci and resolutions. This poster highlights some examples of its use and describes a sampling of the options available for customizing its look and feel. HiGlass is available as an online service, as a docker container or as an open source Github project.

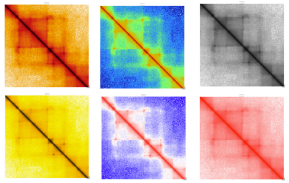
User Interface





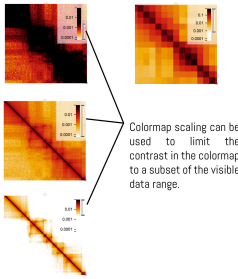
Each track has a set of options that can be accessed via the "configure series" menu. To the left is an example of this menu for the heatmap track.

Colormaps



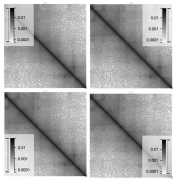
(left to right, top to bottom): fall, jet, greys, custard, bwr, white-to-red

Resolution Limiting

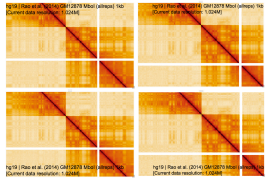


Colormap scaling can be used to limit the contrast in the colormap to a subset of the visible data range.

Colorbar Position



Label Position

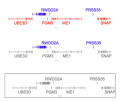


Label Opacity

Track Types

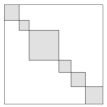
Gene Annotations

Gene annotation tracks show the locations of genes on an assembly. When zoomed in, this track shows the exons and introns as well as the names of the genes.



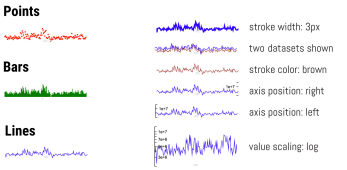
Annotations

Annotations outline a region of interest in the heatmap. They are usually associated with some form of a feature call. We have used this track type to show the locations of TADs and loops called by various programs. When the annotations are chiefly along the diagonal, they can be displayed using the "horizontal rectangular annotations" map.



Points, Bars, Lines

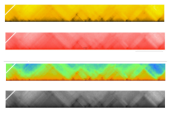
Points, bars and lines are all used to display quantitative 1D genomic information. Because of their similar behavior, they share many common options. For brevity, a sampling of available options is shown for the line track. Similar options can be found for the point and bar tracks as well.



Heatmap

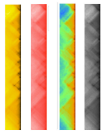
The heatmap is the most prominent track in HiGlass. It is used to display Hi-C contact matrices at varying resolutions. From the choice of color map to the color map scaling, many of the options available are used to customize how this track shows data as well as how the metadata about the track is positioned and shown.

Horizontal Heatmap



The horizontal and vertical heatmaps show 2D matrix data. They are 45 degree rotations of the regular heatmap track. Because they are horizontal tracks, they can only be scrolled along one axis (the diagonal of the heatmap).

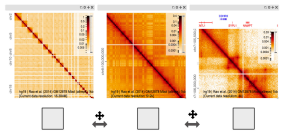
Vertical Heatmap



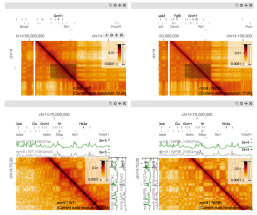
Explore, Compare, and Curate >1000 of 2D Hi-C features

+ BEDPE

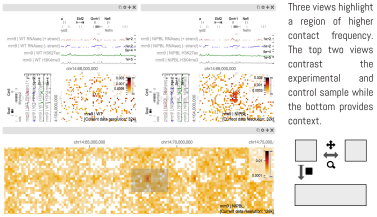
HiGlass Example View Configurations



This configuration shows differences in scale of common Hi-C data features. All three views are centered on the same location. Each zoom level expressed different pattern types.



This configuration has four different views showing data from two samples at two different zoom levels. The two rows are linked by location and zoom. The bottom two views project their viewport onto the views on top for context.



Three views highlight a region of higher contact frequency. The top two views contrast the experimental and control sample while the bottom provides context.

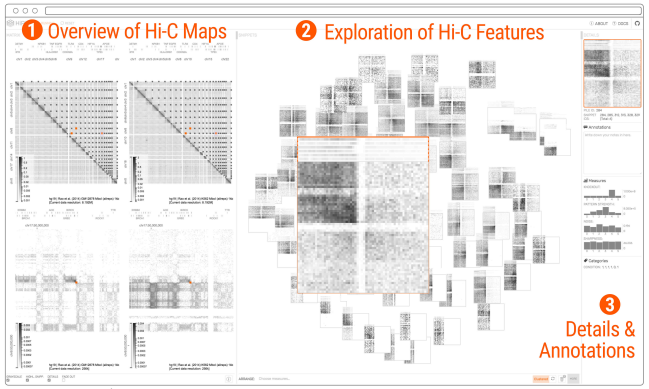
DEMO
http://hipiler.higlass.io

HiPiler: Exploring Many Hi-C Features Through Visual Decomposition.

Fritz Lekschas[1], Benjamin Bach[1], Peter Kerpedjiev[2], Nils Gehlenborg[2], and Hanspeter Pfister[1]

[1]Harvard John A. Paulson School of Engineering and Applied Sciences and [2]Harvard Medical School

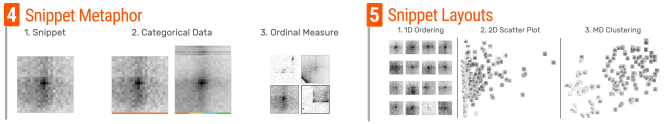
Abstract. HiPiler is a tool for exploring many features in large Hi-C maps. Traditional matrix aggregation or pan-and-zoom interfaces largely fail in supporting dynamic inspection and comparison of thousands of local features (loops, TADs, etc.).

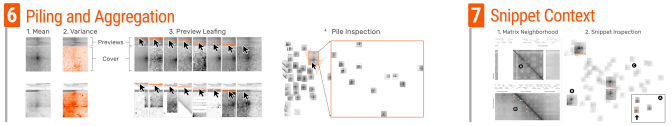


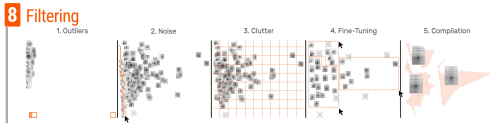
1. Overview of two Hi-C maps (top row) from which features are extracted (highlighted locations). The second row of Hi-C maps shows the detailed neighborhood of one location in Rao et al.'s [3] GM12878 and K562.

2. Telomeric regions are visualized as thumbnails, which can be filtered, arranged, grouped, and aggregated into piles. This example demonstrates multi-dimensional clustering with auto-grouping by spatial distance. An interesting group is enlarged.

3. The sidebar displays detail information of snippets including derived measurements and categories and textual annotation.







4. Snippets are cut-out features from the Hi-C map and visualized as thumbnails. They can display categorical and numerical data.

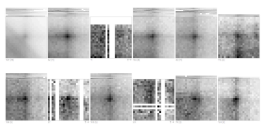
5. Snippets can be ordered and arranged in 1D and 2D along the underlying pattern or by user-defined measurements to identify trends and subgroups.

6. Aggregated snippets are visualized as piles. These piles consist of a cover matrix displaying the average or variance of the snippets and 1D snippet previews. Piles can be inspected in isolation for detailed inspection.

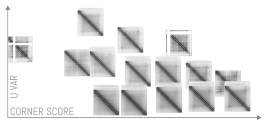
7. The locations of snippets are displayed in the matrix view, which supports detailed exploration of the snippets' context.

8. Manual and semi-automated tools enable rapid filtering and grouping of up to 2 thousand snippets.

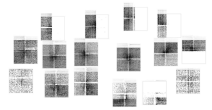
HiPiler Example View Configurations



Overview of loop calls from Rao et al. [3] of chromosome 22 ordered by their distance to the diagonal and grouped by similarity (k-means clustering) for initial exploration.



Analysis of TAD calls from Rao et al. [3] of chromosome 4 along corner score and U var from Arrowhead [3] for performance and parameter estimation of the algorithm. Snippets have been auto-piled by distance and log transformed for clarity.



Variances in telomeric regions in Rao et al. [3] GM12878 and K562 grouped into piles and laid out by pattern-driven clustering. The colors encodes the variance (log-transformed s.d.) within piles. Darker colors indicate higher variance.