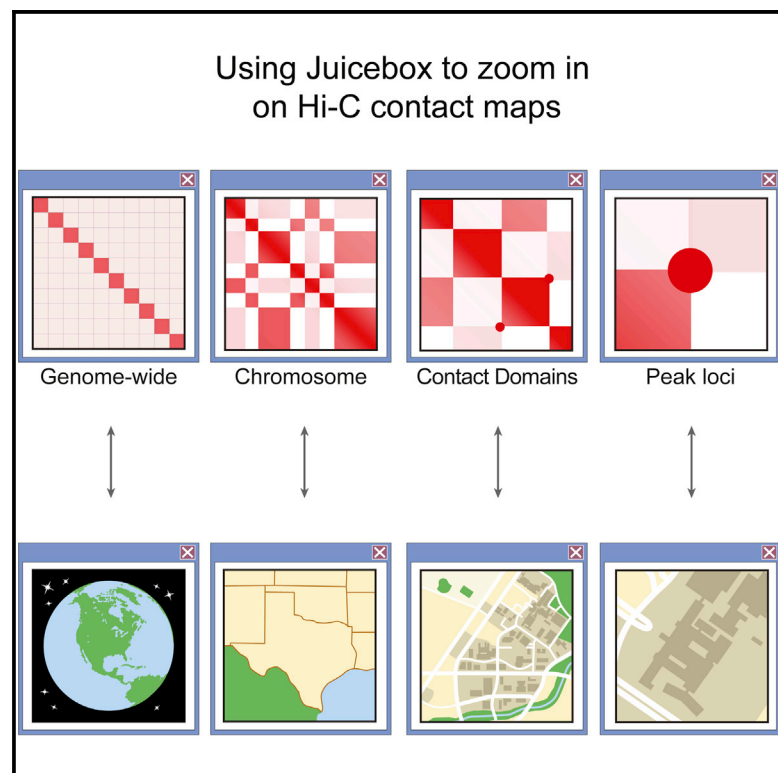# Cell Systems

Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom

Graphical Abstract

Highlights

- Juicebox enables users to explore Hi-C contact maps

- Users can create maps from their own experimental data

- Users can zoom in and out of a contact map in real time

- Maps can be compared to 1D tracks, 2D feature sets, or one another

Authors

Neva C. Durand, James T. Robinson, Muhammad S. Shamim, Ido Machol, Jill P. Mesirov, Eric S. Lander, Erez Lieberman Aiden

Correspondence

erez@erez.com

In Brief

We introduce Juicebox, a tool for exploring contact maps generated using Hi-C and other 3D genome-sequencing technologies. Users can zoom in and out interactively, just as a user of Google Earth might zoom in and out of a geographic map.

CrossMark

CellPress

Juicebox Provides a Visualization System for Hi-C Contact Maps with Unlimited Zoom

Neva C. Durand,[1,2,3,4,8] James T. Robinson,[1,4,8] Muhammad S. Shamim,[1,2,3] Ido Machol,[1,2,3] Jill P. Mesirov,[4]
Eric S. Lander,[4,5,6] and Erez Lieberman Aiden[1,2,3,4,7,*]
[1]The Center for Genome Architecture, Baylor College of Medicine, Houston, TX 77030, USA
[2]Department of Molecular and Human Genetics, Baylor College of Medicine, Houston, TX 77030, USA
[3]Department of Computer Science, Department of Computational and Applied Mathematics, Rice University, Houston, TX 77005, USA
[4]Broad Institute of Harvard and Massachusetts Institute of Technology (MIT), Cambridge, MA 02139, USA
[5]Department of Biology, MIT, Cambridge, MA 02139, USA
[6]Department of Systems Biology, Harvard Medical School, Boston, MA 02115, USA
[7]Center for Theoretical Biological Physics, Rice University, Houston, TX 77030, USA
[8]Co-first author
*Correspondence: erez@erez.com
http://dx.doi.org/10.1016/j.cels.2015.07.012

SUMMARY

Hi-C experiments study how genomes fold in 3D, generating contact maps containing features as small as 20 bp and as large as 200 Mb. Here we introduce Juicebox, a tool for exploring Hi-C and other contact map data. Juicebox allows users to zoom in and out of Hi-C maps interactively, just as a user of Google Earth might zoom in and out of a geographic map. Maps can be compared to one another, or to 1D tracks or 2D feature sets.

Hi-C is a widely employed method for studying how genomes are folded in three dimensions (Rao et al., 2014). In Hi-C, DNA loci that are spatially proximate in a system of interest (such as a cell nucleus or an in vitro DNA preparation) are ligated to one another and sequenced. This results in a Hi-C contact map: a list of pairs of genomic positions that were adjacent to each other in three-dimensional (3D) space. Typically, the pairwise interactions produced by Hi-C experiments are visualized as a heatmap: the linear genome is partitioned into loci of a fixed size, or resolution (e.g., 1 Mb or 1 kb), and each entry in the two-dimensional (2D) heatmap corresponds to the number of contacts observed between a pair of loci during the experiment.

Developing adequate visualizations for Hi-C heatmaps is a challenge, because the size of the meaningful biological features they contain ranges over at least seven orders of magnitude: from loops anchored at 20-bp CTCF sites to territories that extend across 200-Mb chromosomes. At the highest meaningful resolutions, published Hi-C heatmaps contain trillions of entries, and only a tiny portion can be displayed at any given time. At coarser resolutions, more of the map can be shown, but the fine structure can no longer be resolved.

Several Hi-C visualization systems exist. Like a paper atlas, these browsers show the map at a single resolution (typically either 1 Mb or 100 kb), either in its entirety (Servant et al., 2012; Paulsen et al., 2014; Lieberman-Aiden et al., 2009) or very close to the diagonal (Zhou et al., 2013). Users cannot zoom in and out in real time. Also, most of these systems are designed to show specific, previously published datasets, and they do not allow users to visualize their own experiments.

Here we introduce Juicebox, a tool for exploring Hi-C data and contact maps in general (Figure 1). Juicebox allows users to explore Hi-C heatmaps interactively, zooming in and out just as a user of Google Earth might zoom in and out of a geographic map; it integrates many technologies developed for the Integrative Genomics Viewer (Robinson et al., 2011) with a broad ensemble of methods specifically designed for handling 2D contact data. Individual maps can be normalized (corrected for experimental bias), compared to one-dimensional (1D) tracks (such as gene tracks or chromatin immunoprecipitation sequencing [ChIP-seq] data), and compared to 2D feature lists (such as loop and domain annotations). Multiple maps can be browsed side by side simultaneously and compared with one another in various ways, revealing both conservation and variation across cell types and species. Users can create their own heatmaps to explore their own experiments (Durand et al., 2016). Juicebox was an invaluable tool in making biological discoveries across many size scales in our group's recent paper (Rao et al., 2014).

In Juicebox—when using Hi-C maps of adequate depth and quality—increasingly small features can be resolved as the user zooms in. In a genome-wide view, chromosome territories are evident, as are chromosomal rearrangements such as translocations. Clicking on a particular chromosome zooms into its intrachromosomal map, optimized for the user's monitor. The broad compartmentalization of the genome, which manifests as alternating long-range patterns, is typically visible at this resolution. On the X chromosome, two superdomains also may become apparent, partitioning the chromosome. These are accompanied by superloops, bright peaks many megabases away from the diagonal.

Zooming in further can be accomplished by double-clicking, by using the resolution slider, or by drawing a box around a region of interest. At 50-kb resolution, subcompartments can be seen, reflecting finer differences in the long-range contact pattern. At 25-kb resolution, contact domains appear: these are intervals containing loci, which preferentially form contacts with one another and that form squares along the diagonal. Juicebox makes it easy to compare these structures to a large
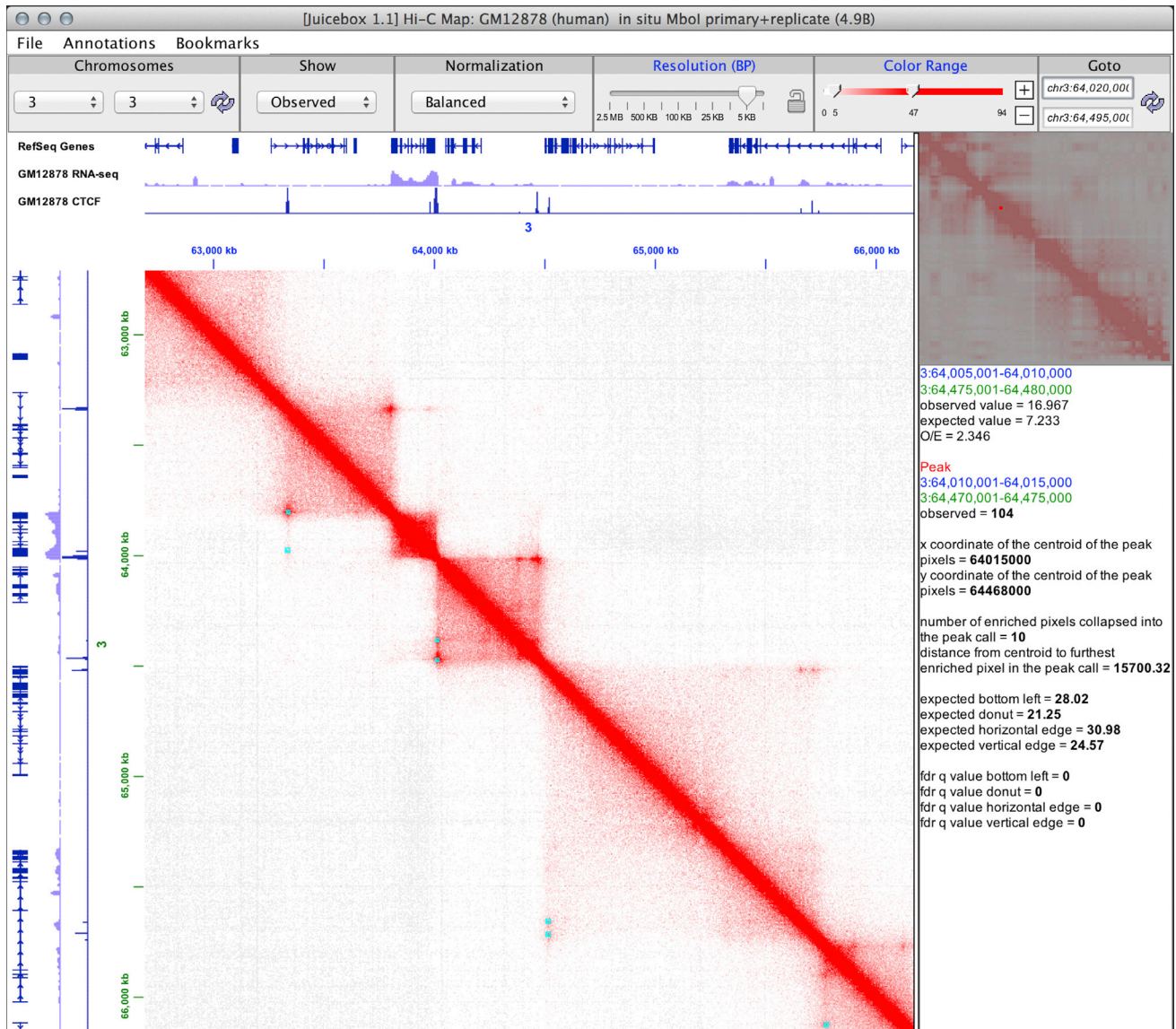
Figure 1. Juicebox Enables Exploration of Contact Maps at Many Resolutions

A screenshot of Juicebox zoomed in to 5-kb resolution on chromosome 3. The toolbar at the top allows users to quickly navigate between different chromosomes, views, normalizations, and resolutions. Users can load one-dimensional tracks and compare them to features seen in the heatmap. Two-dimensional features can be superimposed on the main map. Several peaks are annotated (cyan), each peak indicating the presence of a chromatin loop. A peak in the CTCF track, indicating CTCF binding, is seen at most loop anchors. At the top right, a mini-map shows the whole chromosome at low resolution. Below, hover text shows additional data for one of the highlighted peaks.

number of broad-source chromatin marks at once. We find that the epigenetic marks that decorate particular contact domains correlate strongly with differences in long-range contact pattern.

At 5-kb resolution, chromatin loops are readily seen as bright peaks in which contact frequency is enhanced relative to the local neighborhood. These tend to lie at the corners of contact domains. Finally, at 1 kb, the relationship between the loops and point-source epigenetic tracks can be interrogated. For instance, it becomes clear the chromatin loops are frequently anchored at convergent CTCF motifs, i.e., CTCF motifs that point toward one another (Rao et al., 2014). Juicebox can be used to zoom in much further, although maps with enough data to support such studies do not yet exist.

Juicebox is available as a Java application that can be downloaded and launched via aidenlab.org/juicebox. The code is open source and licensed under the MIT license, available at https://github.com/theaidenlab/Juicebox. Users can explore their own data or examine data from over 15 Hi-C, 5C, and CHIA-PET publications. We also provide the software and test datasets used to review this manuscript at http://dx.doi.org/10.17632/dj4nrsc552.1. We hope that allowing researchers to zoom inward will help studies of chromatin architecture to zoom onward.

REFERENCES

Durand, N.C., Shamim, M.S., Machol, I., Rao, S.S.P., Huntley, M.H., Lander, E.S., and Aiden, E.L. (2016). Juicer Provides a One-Click System for Analyzing Loop-Resolution Hi-C Experiments. Cell Syst. 3, this issue, 95–98.

Lieberman-Aiden, E., van Berkum, N.L., Williams, L., Imakaev, M., Ragoczy, T., Telling, A., Amit, I., Lajoie, B.R., Sabo, P.J., Dorschner, M.O., et al. (2009). Comprehensive mapping of long-range interactions reveals folding principles of the human genome. Science 326, 289–293.

Paulsen, J., Sandve, G.K., Gundersen, S., Lien, T.G., Trengereid, K., and Hovig, E. (2014). HiBrowse: multi-purpose statistical analysis of genome-wide chromatin 3D organization. Bioinformatics 30, 1620–1622.

Rao, S.S.P., Huntley, M.H., Durand, N.C., Stamenova, E.K., Bochkov, I.D., Robinson, J.T., Sanborn, A.L., Machol, I., Omer, A.D., Lander, E.S., and Aiden, E.L. (2014). A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. Cell 159, 1665–1680.

Robinson, J.T., Thorvaldsdóttir, H., Winckler, W., Guttman, M., Lander, E.S., Getz, G., and Mesirov, J.P. (2011). Integrative genomics viewer. Nat. Biotechnol. 29, 24–26.

Servant, N., Lajoie, B.R., Nora, E.P., Giorgetti, L., Chen, C.-J., Heard, E., Dekker, J., and Barillot, E. (2012). HiTC: exploration of high-throughput 'C' experiments. Bioinformatics 28, 2843–2844.

Zhou, X., Lowdon, R.F., Li, D., Lawson, H.A., Madden, P.A.F., Costello, J.F., and Wang, T. (2013). Exploring long-range genome interactions using the WashU Epigenome Browser. Nat. Methods 10, 375–376.