Project Acronym: MuG

Project title: Multi-Scale Complex Genomics (MuG)

Call: H2020-EINFRA-2015-1

Topic: EINFRA-9-2015

Project Number: 676556

Project Coordinator: Institute for Research in Biomedicine (IRB Barcelona)

Project start date: 1/11/2015

Duration: 36 months

Deliverable 3.3: A browser-track that implements and connects all the 2D data from a genome or a genomic domain

Lead beneficiary: Institute for Research in Biomedicine (IRB Barcelona)

Dissemination level: PUBLIC

Due date: 31/10/2016

Actual submission date: 15/11/2016

This project has received funding from the European Union's Horizon 2020 research and innovation programme under grant agreement No 676556.

Document history

Version	Contributor(s)	Partner	Date	Comments
0.1	David Castillo	CNAG-CRG	30/10/2016	First Draft
0.2	Marc A. Martí-Renom	CNAG-CRG	4/11/2016	Second Draft
0.3	Mike Goodstadt	CNAG-CRG	15/11/2016	Updated TADkit browser URL
1.0			15/11/2016	Final version, approved by Technical Board and Supervisory Board

Table of Contents

Executive summary

Interactions within a genome reveal its form and function and are detected by Hi-C experiments plotted as 2D heat maps or matrices. However, interactions between distant genomic locations do not fit in traditional chromosome-specific non-global genome browsers. Also matrix height impedes legibility when stacked with linear browser tracks. This document describes the development of a component of the MuG visualization tool TADkit (http://sgt.cnag.cat/3dg/mug/tadkit/), for coherent visualization of 2D interaction matrices.

1 INTRODUCTION

Recent research using Hi-C experiments to detect interactions within the genome has brought a new understanding of the role of structure in genome function. However, it also brings new challenges in visualization of the resulting data for analysis [1]. The interactions can be between distant genomic locations whereas classic genome browsers generate feature-specific small-scale-rich views. Also, the complete set of interactions are usually plotted all-to-all on 2D heat maps or matrices [Fig.1] which occupy significant vertical layout space and so cannot be comfortably represented within traditional stacking of browser tracks [Fig.2].

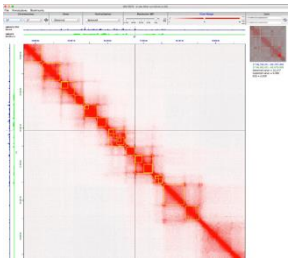


Figure 1. Juicebox [2]

Nowadays many HiC matrix software applications and viewers exist although only a few of them are harmonized with classical 1D browsers [2]. Typically half of the matrix is displayed with the diagonal aligned to linear tracks but these lack synchronized matrix and track navigation and annotation [3][4].

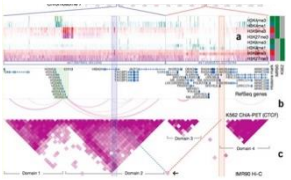
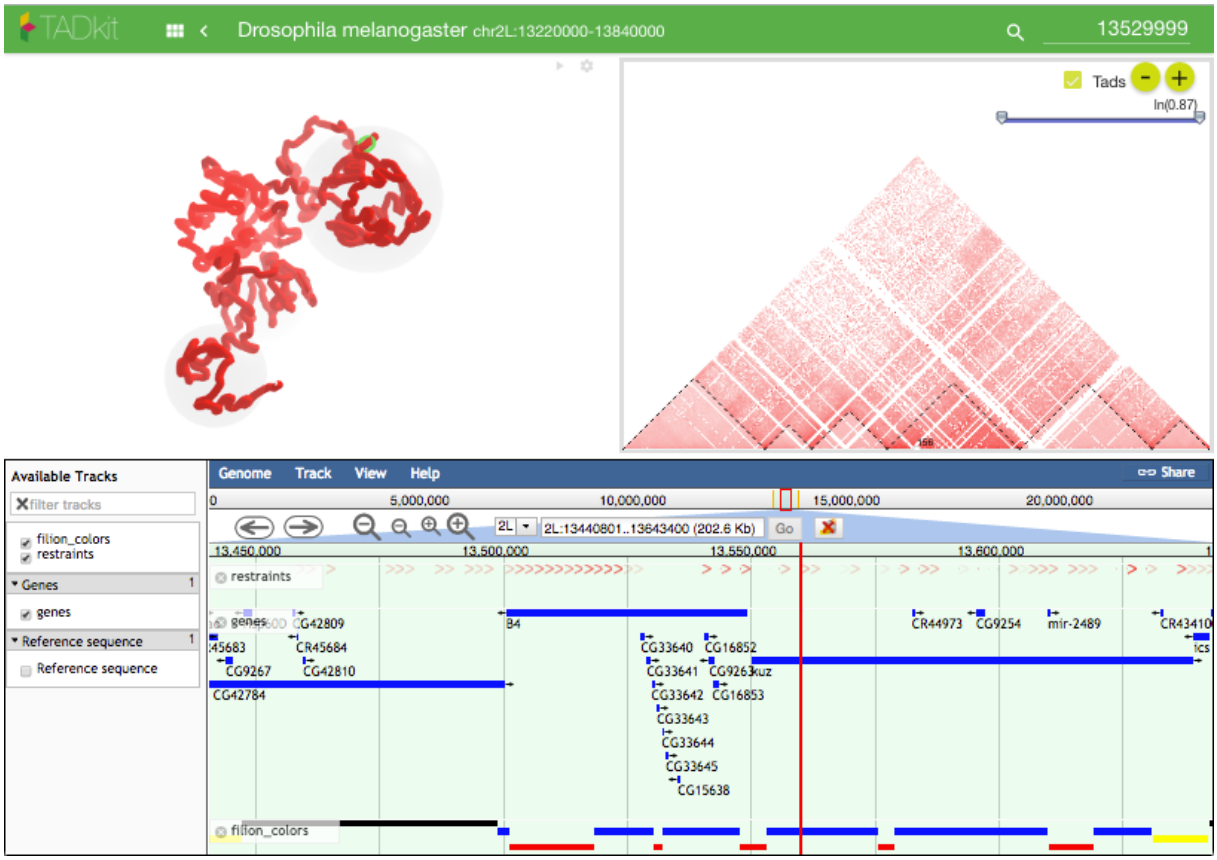


Figure 2 WashU Browser [3]

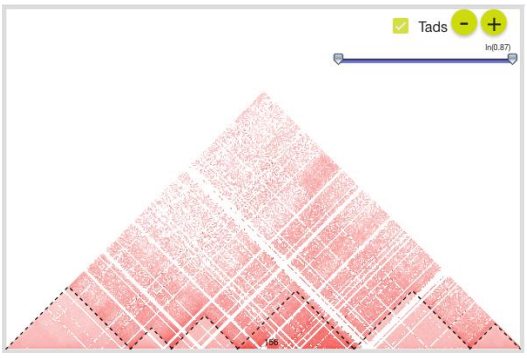


A 2D component has been developed for TADkit where Hi-C data is displayed as interactive interaction matrices coordinated with classic genomic tracks. See http://sgt.cnag.cat/3dg/mug/tadkit/

2 2D COMPONENT IN TADKIT

2.1 Design

The 2D panel in TADkit displays half of the symmetric interaction matrix rotated as to have the diagonal in the horizontal axis. As mentioned in the introduction this kind of representation is common in many Hi-C viewers and allows the end-user to easily relate the 2D information with the classical 1D tracks which are also displayed along the horizontal axis.



The panel is a html canvas component that has been placed in the upper-right side of the screen. The view of the HiC matrix occupies a square space which is difficult to align with 1D tracks.

2.2 Features

The matrix can be moved along its horizontal base and zoomed in or out as desired. These move and zoom features helps to easily inspect the information contained.

Hi-C interaction matrices are represented in logarithmic scale. By clicking on any point of the matrix the value of the frequency of interaction is shown.

A slider has been also implemented to filter interactions by its values. The filtering of the data in the matrix is important to highlight significant interactions or depleted regions.

Another important development to help in the analysis of the scientific data is the identification of the Topologically Associating Domains (TADs) marked with dashed lines. TADs will be identified in the matrix if the source data includes the information about the TAD boundaries.

3 CONNECTING 1D AND 2D COMPONENTS

There is interconnection between the different components of TADkit. The 2D matrix displayed data region is highlighted in the 1D browser and the navigation in the linear genomic tracks is also identified as a moving point in the 2D component.

A further integration consists in the identification of two interacting loci. By clicking on any point of the 2D matrix two marks are placed in the genomic positions of the interacting loci.

4 CONCLUSIONS AND FUTURE PERSPECTIVE

A 2D component has been developed and integrated in TADkit. This component is effectively connected with the 1D tracks of the browser. Both the component and the connection with the other components will be improved based on the feedback from the MuG community. It might be interesting, for instance, for the component to highlight significant loops based on certain threshold. The development of TADkit is an ongoing process that will be enriched by the input of the scientific community and specially the MuG pilot projects.

5 REFERENCES

1. Dekker, J., Marti-Renom, M. A., & Mirny, L. A. (2013). Exploring the three-dimensional organization of genomes: interpreting chromatin interaction data. Nat Rev Genet, 14(6), 390–403.

2. Neva C. Durand*, James T. Robinson*, Muhammad S. Shamim, Ido Machol, Jill P. Mesirov, Eric S. Lander, and Erez Lieberman Aiden. "Juicebox provides a visualization system for Hi-C contact maps with unlimited zoom." Cell Systems 3(1), 2016.

3. Zhou, X., Lowdon, R. F., Li, D., Lawson, H. A., Madden, P. A. F., Costello, J. F., & Wang, T. (2013). Exploring long-range genome interactions using the WashU Epigenome Browser. Nat Meth, 10(5), 375–376.

4. YUE lab 3D Genome Browser: http://www.3dgenome.org