



Multifiscale Complex Genomics



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 3.6: A 3D Genome browser connecting all the 1D-2D-3D tracks. The browser will allow different renderings of the upload coordinates

Lead beneficiary: Fundació Centre de Regulació Genòmica (CRG)

Dissemination level: PUBLIC

Due date: 31/10/2017

Actual submission date: 10/11/2017

Copyright© 2015-2018 The partners of the MuG Consortium



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	22/10/2017	First Draft
0.2	Mike Goodstadt	CNAG-CRG	25/10/2017	Second Draft
0.3	Marc A. Marti-Renom	CNAG-CRG	27/10/2017	Third Draft
1.0			10/11/2017	Approved by Supervisory Board





Table of Contents

1	INTRODUCTION	5
2	GENOME BROWSER CONNECTING ALL THE 1D-2D-3D TRACKS	5
3	3D VIEWS.....	7
3.1	PARTICLES VIEW	7
3.2	CHROMATIN VIEW.....	8
3.3	CLUSTER VIEW	8
4	REFERENCES.....	10





Executive summary

The visualization of 3D conformations of chromatin gives an innovative and new perspective in the understanding of biological processes. Those structures are not by far simple points in space but contain rich information about the region of the genome from which they are derived. It is therefore important to provide to the scientist different views of the information that will help him in the analysis of aspects like variability, expression or accessibility of the chromatin. TADkit in the MuG VRE (<http://vre.multiscalegenomics.eu/>) include different possibilities for the rendering of 3D structures while maintaining an integrated 1D, 2D and 3D workspace.



1 INTRODUCTION

The MuG Virtual Research Environment (VRE) includes TADbit (<http://www.3DGenomes.org/tadbit>), a complete Python library to analyze, model and explore 3C-based data [1]. TADbit computes and cluster models that best fit the interaction matrix of a defined region of the genome obtained in the 3C-based experiment and saves the resulting information in a JSON dataset that can be viewed in TADkit, a unified workspace which connects the 3D data with other data in that genomic domain.

The dataset of chromatin structures contains valuable information about the biological insights of the experiment, information that is difficult to reveal with the simple representation of the 3D coordinates. It is therefore important to provide different views or renderings of the information contained in the dataset. TADkit allows the visualization of the 3D structures in three different and combinative views: particles view, chromatin view and cluster view.

Each model in the dataset is a set of consecutive three-dimensional space points which are joined to represent the chromatin strand. Models in the dataset are clustered by structural similarity in TADbit. Each cluster represents then a particular state or subpopulation of the whole ensemble, a family of possible 3D structures equally compatible with the experimental data.

When a dataset is opened in TADkit, the user is asked to select which cluster is about to inspect and he/she is conducted to the visualization of the centroid model of the selected cluster. The centroid is the existing model in the cluster which more resembles the all other models of the cluster. The views can be activated or deactivated by accessing the configuration options in the upper right part of the 3D panel in TADkit [Figure 1].

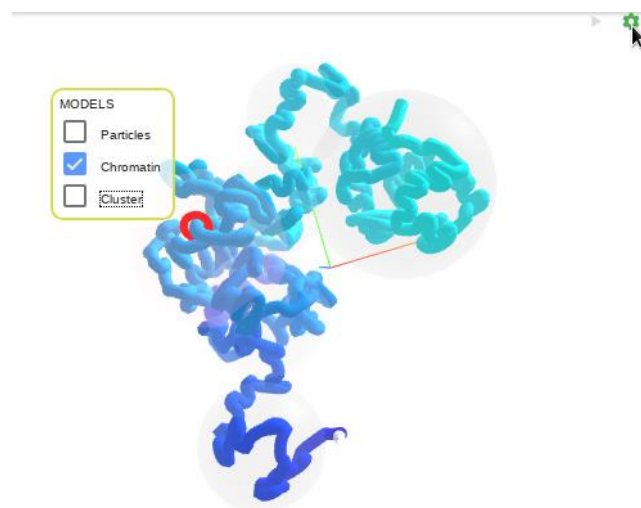


Figure 1: Configuration options in the 3D panel of TADkit.

2 Genome browser connecting all the 1D-2D-3D tracks

TADkit was integrated in the VRE infrastructure with the connection of 1D, 2D and 3D tracks in M18 as described in Deliverable 3.4. The continuous development of the tool has led to the inclusion of a fourth panel devoted to the display of multiple types of information.

The panel is placed in the bottom left part of the workspace [Figure 2] and is organized in different categorical tabs. At the time of writing the panel is composed by two tabs: Interactions and Dataset Settings.

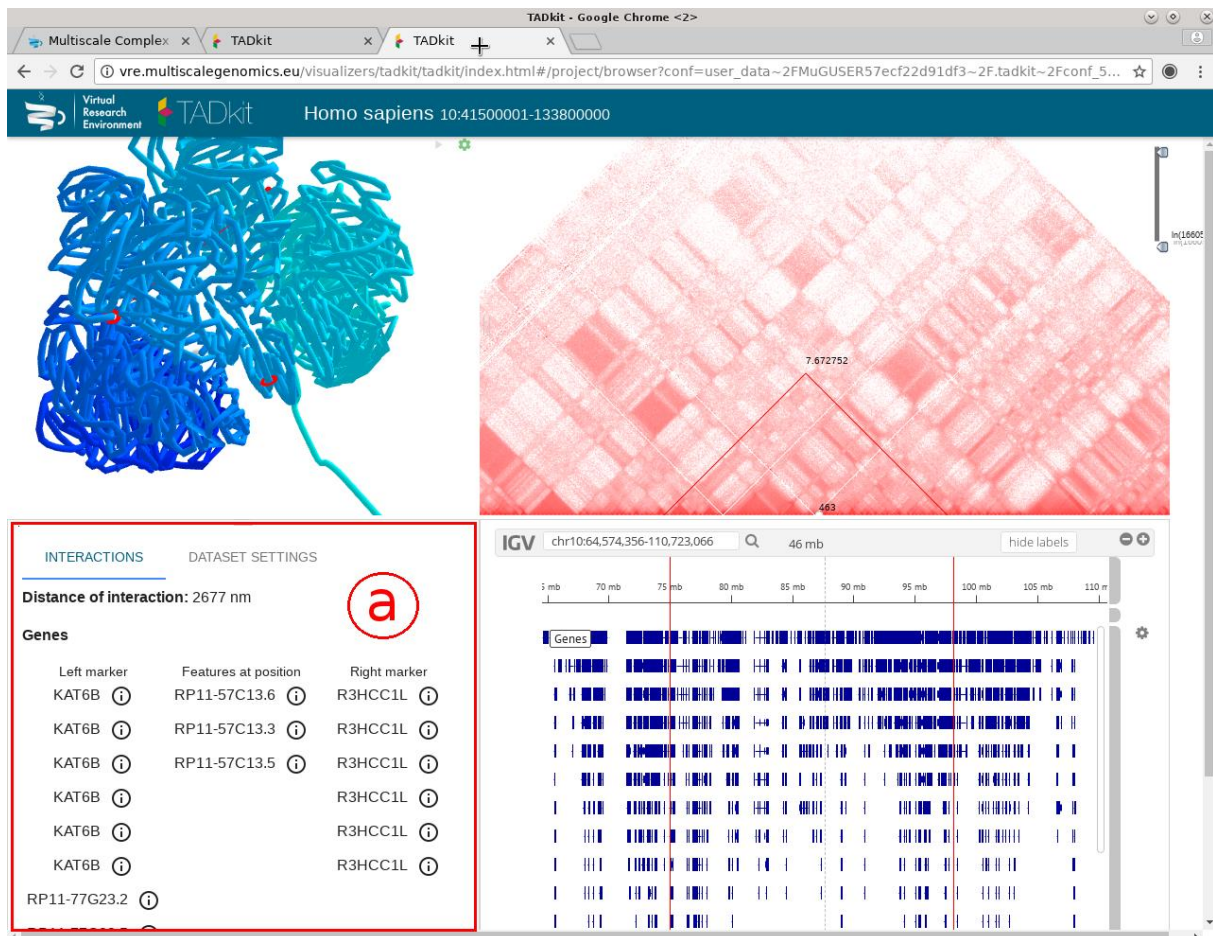


Figure 2: (a) Information panel of TADkit

The Interactions tab [Figure 2] displays the features from the existing 1D tracks located in the current genomic position marked by the gray vertical dashed line. It gets further populated when the user clicks in the 2D matrix to select an interaction between two genomic loci. The panel will then display the distance between those two loci in the model rendered in the 3D panel. It will as well list the features from the existing 1D tracks located in those two genomic positions.

The Dataset Settings tab [Figure 3] gathers information about the dataset file loaded in TADkit. It contains information such as the biological species, the represented genomic region, the resolution of the 3C-based interaction matrix and the number of bins of the model.

INTERACTIONS

DATASET SETTINGS

Sample TADbit data


Species:	Homo sapiens
Region:	10:41500001-133800000
UUID:	b5de94af-3f42-5ab4-a812- fb195e8c6389
Resolution:	100000
Bins:	924
Chromatin radius:	5 nm
Chromatin radius scale:	 10

Figure 3: Dataset Settings tab in the information panel of TADkit

3 3D Views

3.1 Particles view

The particles view represents the chromatin model as a set of consecutive spheres centered on the 3D points of the model coordinates [Figure 4]. The radius of the spheres is proportional to the resolution of the dataset.

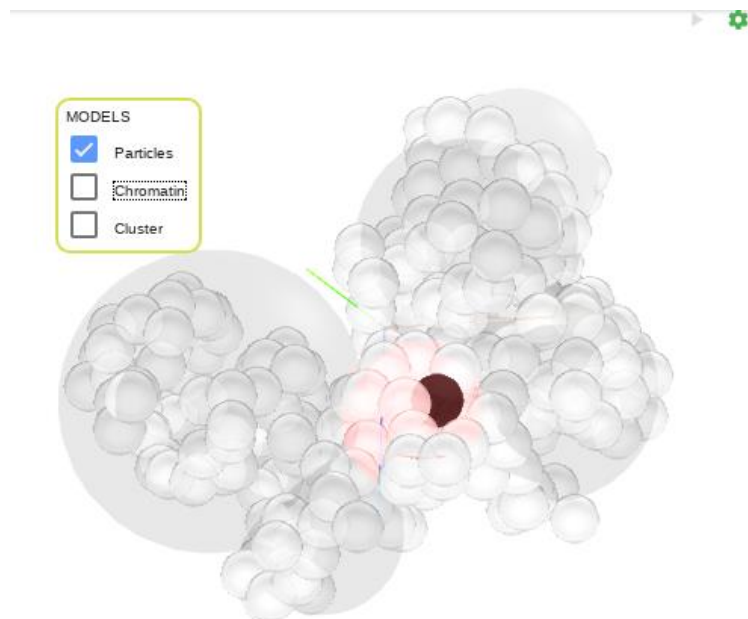


Figure 4: Particles view of the 3D panel of TADkit

The current genomic position in the navigation of the linear tracks is identified by coloring in black the sphere of the current 3D coordinate.

3.2 Chromatin View

The chromatin view represents the model as a tubular strand where the 3D points of the model are surrounded by a limited number of segments. The segments join the consecutive points to simulate the continuity of the chromatin strand. Depending on the resolution used to compute the model the strand is given an appropriate thickness to approximately match chromatin width [Figure 5].

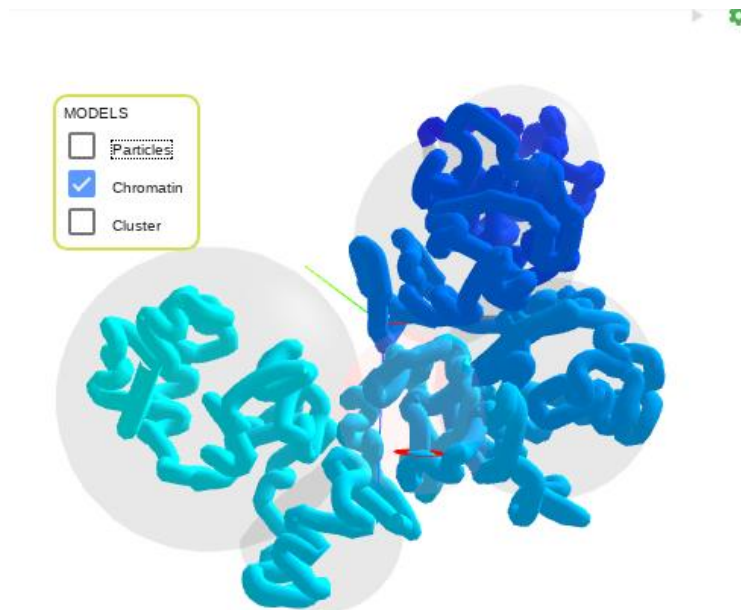


Figure 5: Chromatin view of the 3D panel of TADkit

This representation of the chromatin allows the coloring of the segments that form the strand. A menu option in the 1D tracks can be used to color (overlay) the model depending on the information of the selected track. If the feature includes a color attribute, that color will be overlaid in the chromatin strand through the genomic region occupied by the feature. If the feature includes a score or value attribute; that score or value will be used to create a blue gradient color table used to overlay the strand.

3.3 Cluster View

The previous views are used to visualize single models in the dataset, usually the centroid. On the other hand, the cluster view represents at the same time the set of models belonging to an individual cluster.

The models of the selected cluster are rendered as thin straight lines connecting the consecutive three-dimensional points [Figure 6]. Since clusters can contain large numbers of models, simple lines have been chosen to represent them and minimize the use of memory resources in the 3D panel.

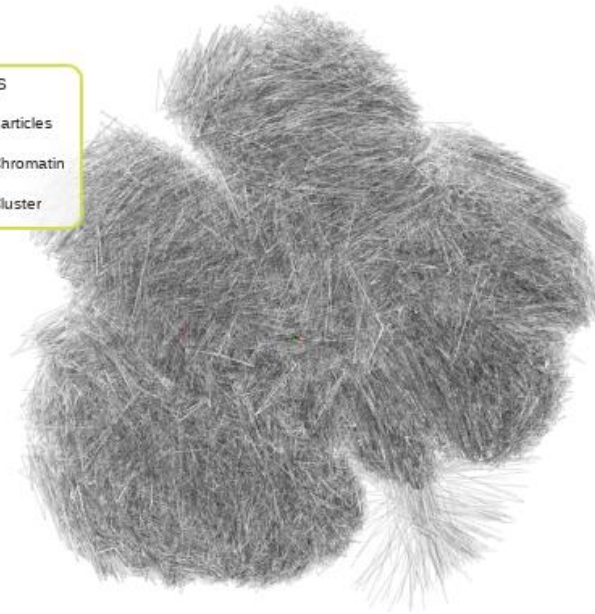
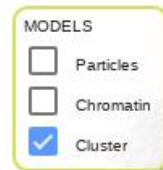


Figure 6: Cluster view of the 3D panel of TADkit

The visualization of the ensemble of models gives valuable information about the variability of the cluster and the occupancy of the structure as a dynamic strand.



4 REFERENCES

1. Serra, F., et al., *Automatic analysis and 3D-modelling of Hi-C data using TADbit reveals structural features of the fly chromatin colors*. PLoS Comput Biol, 2017. **13**(7): p. e1005665.

