Genome3D: A Viewer-Model Framework for Integrating and Visualizing Multi-Scale Epigenomic Information Within a Three-Dimensional Genome

William Arndt, Thomas M. Asbury, W. Jim Zheng
Bioinformatics Division
Department of Biochemistry and Molecular Biology
Medical University of South Carolina
Charleston, SC
arndtw@musc.edu

Matt Mitman, Jijun Tang
Department of Computer Science Engineering
University of South Carolina
Columbia, SC
jtang@cec.sc.edu

Abstract—New technologies permit the measurement of many types of genomic and epigenomic information at scales ranging from the atomic to nuclear. Much of this new data is structural in nature and challenging to coordinate with existing data formats. There is an increasing need to integrate and visualize disparate data sets in order to reveal structural relationships not apparent when viewing these data formats separately.

We have developed the Genome3D software package in order to integrate and display epigenomic data within the context of a three-dimensional physical model of the human genome. To our knowledge, this is the first such tool developed to visualize human genome in three dimensions. Here we describe the major features of Genome3D and discuss our multi-scale data framework using a representative basic physical model.

Additionally, utilities to generate demonstration models of common chromosome interaction scenarios derived from High-C or CHIA-PET analysis are under development. These models are compatible with existing annotation data formats and can demonstrate common situations where one or more chromosomes are arranged such that two or more interaction sites appear in close proximity. These models can be used to directly create publication images without resorting to intermediate use of image editing software.

Availability:
http://genombioinfo.musc.edu/Genome3D/Index.html

I. BACKGROUND

Recent advances in high throughput technologies have granted easy access to genetic sequence data, but this data alone is insufficient to model gene expression and behavior. A large number of epigenetic characteristics can act in tandem with DNA to regulate genes, differentiate cell types, or cause disease. These epigenetic characteristics have widely varying scales and influences: from the simplest methylation of a single nucleotide, the various possible modifications of histone proteins affecting binding sites and chromatin fiber structure, to the largest view of the three dimensional arrangement and packing of chromatin fiber and chromosomes.

With the influx of new detail about the non-sequence characteristics of the genome, new techniques will be required to visualize and model the full extent of genomic interactions and function. Genome browsers, such as the USCS Genome Database Browser [2], are specifically aimed at viewing primary sequence information. Although supplemental information can easily be annotated via new tracks, representing structural hierarchies and interactions is quite difficult, particularly across non-contiguous genomic segments [3]. In order to make sense of the vast amounts of genetic and epigenetic information available, it must be possible to present this information in a manner that unifies the various types of data in proper structure, scale, and context.

II. GENOME3D VIEWER

Genome3D is a software application designed to visualize genetic and epigenomic information in a combined framework. This viewer arranges the genome in a hierarchy from individual atoms up to the arrangement of chromosomes and in this way is able to manage the entire physical genome. Our model framework is flexible and adaptable in order to handle more precise structural information as new details are revealed by advancing research.

A hierarchy of data sources are used to display a model in Genome3D and, when possible, arrangements are determined by data from other levels of the hierarchy in order to conserve storage space. The primary source of information used to create the display is a sequence of histone complex positions and orientations. From this list, the positions of nucleotides and atoms are calculated automatically. FASTA sequence files can be loaded and displayed on the corresponding DNA fragments. A chromatin fiber strand is drawn using a sequence of 3D control points chosen to position the curve as desired. Finally, annotations representing epigenomic data can be loaded from common data formats such as BED and WIG [2] and their visual cues added to the 3D structure of the genome. This structure is displayed with a fully controllable camera, has three distinct zoom levels to properly scale each layer of detail, and can output both images, ray tracing scenes in PovRay format, or PDB files.
III. Visualizing Common Gene Interactions

Utilities are under development which can automatically create genome models to demonstrate common interaction scenarios. In the first example, two interaction sites occur on a single chromosome such that they are pulled together and the fiber is curved into a loop structure. A user specifies the location and size of these interaction sites and the corresponding fiber and nuclear core particle models are generated automatically. These models are loaded into Genome3D which can then directly accept and display annotations for viewing. In the second situation, two or more chromosomes contain interaction sites to be displayed in close proximity.

The user specifies the number of chromosomes and the length of segment to be displayed for each. The appropriate models are created, loaded into Genome3D, and ready to accept annotations.

Figure 2 shows an example image from a publication on the subject of the structure of chromatin domains. The image was created by hand with image editing software. Figure 3 shows an image produced by the Genome3D utility for single chromosome interactions. The fiber strand was created based on user specified information about the relative size and position of the primary interaction site and annotations were directly applied to show the other features of the image.
Figure 4 shows the model based upon the same image but viewed on the NCP scale.

IV. A 3D Genome Model

To illustrate the capability of Genome3D to integrate and examine data of appropriate scales, we constructed an elementary model of the physical genome by generating a random walk fiber strand guided within known chromosome territories and with appropriate fiber density. Precise knowledge of the physical genome layout is largely unknown at present; however, this lack of reliable information is secondary to visualization approach which provides a springboard for future model development. Current technologies are making significant progress toward capturing detailed chromosome conformation information. Our multi-scale framework permits description on multiple scales so it can easily incorporate disparate new sources of structural folding information, such as fractal globule behaviour, space filling curve models, or data derived for future experimental studies. The Genome3D viewer, decoupled from the genome model, can be used to view any model which can be mapped to simple cartesian coordinates.

Building a 3D model of a complete physical genome is a non-trivial task. The structure and organization at a physical level is dynamic and heavily influenced by local and global constraints. A typical experiment may provide new data at a specific resolution or portion of the genome, and the integration of these data construct a coherent multi-resolution model is challenging. For example, an experiment may measure local chromatin structure around a transcription site. This structure can be expressed as a collection of DNA strands, NCPs, and perhaps lower resolution 30nm chromatin fibers. Our data formats are flexible enough to allow partial integration of this information, when the larger global structure is undetermined, or inferred by more global stochastic measurements from other experiments. Combining such data across resolutions is often difficult, but establishing data formats and visualization tools provide a framework that may simplify the integration process.

REFERENCES

[1] C.A. Espinoza and Bing Ren, "Mapping Higher Order Structure of Chromatin Domains," Nature Genetics, vol. 43, no. 7, pp. 615-616, 2011.

[2] Karolchik D, Kuhn RM et. al, "The UCSC Genome Browser Database: 2008 Update," Nucleic Acids Research, 2008.

[3] Dekker J, "Gene Regulation in the Third Dimension," Science, 2008, 319(5871):1793-1794.

Figure 2. An example figure showing two interaction sites which have formed a loop in a single chromosome. Figure taken from work by C.A. Espinoza [1].

Figure 3. A chromosome fiber model with annotations applied in order to duplicate the example figure.

Figure 4. A chromosome nucleotide model with annotations applied to replicate the situation described by the example figure.