The GLOBE 3D Genome Platform
A New Paper Tool for Systems Biological/Medical Data Integration

Michael Lesnussa 1, Frank N. Kapper 1, Hubert B. Eussen 2, Petros Kolovos 1, Frank G. Grosveld 3 & Tobias A. Knoch 4

in cooperation with the virtual EpigenSys laboratories of K. Rippel 1, P. R. Cook 1, G. Längst 4 & G. Wedemann 1

The systems biological/medical combination of genome sequence and structure, its annotation and experimental data in an accessible and comprehensible way is a major challenge. Increasingly there is a large number of extremely divergent data sets: the sequence itself, genes, regulatory regions, various forms of recurring sequence features and clone sets etc. Currently, one possibility to represent this information in a visual form - and thus to reveal its scientific meaning - is to use genome browsers such as “Ensembl” or the “UCSC Genome Browser”. These browsers have been beneficial in the understanding of the complex organization of genomes. However, there are also limitations concerning their data for these presentation, standardized input and data base accessibility. One customizability by a remote user with special requirements is difficult. Here we show successfully with the GLOBE 3D Genome Platform ways to visualize multi-dimensional data sets from various sources in an easily accessible way. This allows the integration of these data sets into a single holistic virtual display system giving a systems biological/medical oriented view of genomes advancing basic research, diagnostics and new treatments.

Introduction

The platform allows the mapping of chemical and experimental data tracks projected onto metaphase chromosomes (Fig. 1). The general track and every single track element layout is customizable e.g. in position, shape and color. The viewer allows to visualize in principle an unlimited number of elements.

Features

Flexible Customizable Intuitive Navigation
Real-Time Interaction & Analysis
Dynamical Resolution & Arrangement
Extensive Large & Multi-Dimensional Data
Bridge ALL Scales from Sequence to Morphology

Multi-Mapping

The platform allows the mapping of chemical and experimental data tracks projected onto metaphase chromosomes (Fig. 1). The general track and every single track element layout is customizable e.g. in position, shape and color. The viewer allows to visualize in principle an unlimited number of elements.

Inter-Relations

In addition to the simultaneous mapping on one chromosomes, the platform allows the analysis of inter-chromosomal relationships based either on an external input (Fig. 2) or internal correlation analysis (Fig. 1B). Enhanced downstream interaction to regulated SNPs or symptoms to duplications or gene families to breakpoints.

Intra-Relations

Using the dynamic scaling range of the intra-chromosomal relationships can be studied in detail in the track mapping (Fig. 3) concerning basic research, diagnostics and medicine. Assembly over a wide range of data sets enables visualizing on various genome levels to scale-free insights.

Resolution Scale

The platform has a large dynamic range in the size and correlation of all chromosomes in one display from whole genome structure to single gene level (Fig. 5). This new environment creates extra new possibilities for understanding genome organization.

Data Tracks

Syndrome Break Points Duplication Repeats
Genes /SNP DNA Chromosomes Chromatin Histones

Conclusion

The GLOBE 3D Genome Platform presents a novel approach to visualize and analyze the multi-dimensional aspects of a genome in a novel way. In combination with a high-resolution and a computing grid and linking array in parallel, an environment with online and offline processing has been created. This unique new presentation allows the visualization of the multi-dimensional and multi-relational properties of genomes, which is necessary for advanced diagnostic and experimental medicine.

Structure

There are several physical levels of genetic information storage, e.g. DNA, chromatin and chromosomes. This information drives all cell processes and is of critical importance for genome function. The platform visualizes 3D genomic structures and to project storage, e.g. DNA, chromatin and chromosomes. The platform visualizes 3D genomic structures and to project storage, e.g. DNA, chromatin and chromosomes. The platform visualizes 3D genomic structures and to project storage, e.g. DNA, chromatin and chromosomes.

Fig. 1: Complete merged clone set (UCSC, NCBI, Ensemble) of chr. 10. Colors represent association with duplicate regions.

Fig. 2: Various data tracks.

Fig. 3: The basic data tracks.

Fig. 4: Various data tracks.

Fig. 5: Intra-chromosomal duplicons (Eichler et al.) compared to syndromes (blue/orange), fluorescent in-situ (fis) (green), and our defined terminally (red) in the chr 21q11 region.

Fig. 6: Dynamic zoom into the level of the DNA.

Fig. 7: Background image: Multi-chromosomal relation between the duplications of chr. 15 to all other chromosomes. Colors are in Fig. 5.

Fig. 8: Multi-chromosomal relation, view between duplicon regions between chr. 15 & 21. Colors depict spreading degree.
The GLOBE 3D Genome Platform – A New Virtual Paper Tool for Systems Biological/Medical Data Integration

Lesnussa, M., Kepper, N., Eussen, H. B., Kolovos, P., Grosveld, F. G. & Knoch, T. A.

EraSysBio+ Mid-Term Conference, Kaiserwasser Hotel, Vienna, Austria, 13th - 15th September, 2011.

Abstract

Genomes are tremendous co-evolutionary holistic systems for molecular storage, processing and fabrication of information. Their system-biological complexity remains, however, still largely mysterious, despite the huge advances in the understanding of the general sequential, three-dimensional and regulatory organization. With the development of the GLOBE 3D Genome Platform we have created a completely novel grid based virtual “paper” tool and in fact the first systems biological/medical genome browser integrating the holistic complexity of genomes in a single easy comprehensible way, which is used in WP1-5. Based on a detailed study of biophysical and IT requirements, every architectural level from sequence to morphology of one or several genomes can be approached in a real (WP1-3) and in a simulated symbolic representation (WP4) simultaneously and navigated by continuous scale-free zooming within a three-dimensional OpenGL and grid driven environment. In principle several multi-dimensional data sets can be visualized, customized in terms of arrangement, shape, colour, and texture etc. as well as accessed and annotated individually or in groups using internal or external data bases/facilities. Hence, the GLOBE 3D Genome Platform is an example of a grid based approach towards a virtual holistic desktop for system biological/medical genomic work combining the three fundamental distributed resources: i) visual data representation, ii) data access and management, and iii) data analysis and creation.

Corresponding author email contact: TA.Knoch@taknoch.org

Keywords:

Genome, genomics, genome organization, genome architecture, structural sequencing, architectural sequencing, systems genomics, coevolution, holistic genetics, genome mechanics, genome statistical mechanics, genomic uncertainty principle, multilism genotype-phenotype, genome function, genetics, gene regulation, replication, transcription, repair, homologous recombination, simultaneous co-transfection, cell division, mitosis, metaphase, interphase, cell nucleus, nuclear structure, nuclear organization, chromatin density distribution, nuclear morphology, chromosome territories, subchromosomal domains, chromatin loop aggregates, chromatin rosettes, chromatin loops, chromatin quasi fibre, chromatin density, persistence length, spatial distance measurement, histones, H1.0, H2A, H2B, H3, H4, mH2A1.2, DNA sequence, complete sequenced genomes, molecular transport, obstructed diffusion, anomalous diffusion, percolation, long-range correlations, fractal analysis, scaling analysis, exact yard-stick dimension, box-counting dimension, lacunarity dimension, local nuclear dimension, nuclear diffuseness, parallel super computing, grid computing, volunteer computing, polymer model, analytic mathematical model, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization (FISH), targeted chromatin capture (T2C) confocal laser scanning microscopy, fluorescence correlation spectroscopy, spatial precision distance microscopy, super-resolution microscopy, two dimensional fluorescence correlations spectroscopy (2D-FCS) auto-fluorescent proteins, CFP, GFP, YFP, DsRed, fusion protein, in vivo labelling, information browser, visual data base access, holistic viewing system, integrative data management, extreme visualization, three-dimensional virtual environment, virtual paper tool.
Literature References


resolution high throughput method to detect genomic interactions and regulatory elements. *Epigenetics & Chromatin* 7:10, 1-17, 2014.


