A New Holistic Genome Viewer for Molecular Cytogenetics

Bert Eussen 1,2, Michael J. Moorhouse 1,2, Michael Lesnussa 1,4, Maarten Muetgeert 1,4 and Tobias A. Knoch 1,4

In cooperation with A. de Klein 1, P. v.d. Spek 1, F. G. Grosveld 1

The GLOBE-Consortium
headed by B. Eussen, M. Moorhouse and Tobias A. Knoch

Department of Clinical Genetics, Department of Bioinformatics and Biophysical Genomics Group in the Department of Cell Biology
Erasmus Medical Center, Dr. Molewaterplein 50, NL-3015 GE Rotterdam, The Netherlands
http://www.arasmuerc.nl/globe-consortium or globe@erasmuerc.nl

Introduction

The 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 reoccurring 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 focus on linear presentation, standardized input and data bank accessibility. Also comprehensibility by a remote user with special requirements is difficult. The GLOBE-Consortium is developing 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 display system giving a biological oriented view of genomes and advancing basic research, diagnostics and new treatments.

Multi-Mapping

The viewer allows the mapping of classical, and experimental data tracks projected onto metaphase chromosomes simultaneously (Fig. 1). The general track as well as every single track element layout is customizable e.g. chromosome bands and elements can be adjusted and visualized in order to present an unlimited number of elements.

Data Tracks

Syndrome
Break Points
Duplicon
Repeat Regions
Epigenetics
Genes /SNP
DNA
Chromosome
Intergen Band
Chromatin Loops
Chromatin Fiber
Nucleus

Fig. 1: Complete merged clone set (UCSC, NCBI, Ensembl) of chr. 15; colours represent association with duplication regions.

Intra-Relations

Using the dynamic scaling range of the intra-chromosomal structure, all relationships can be visualized to support diagnostics and treatments. Always can be projected, rotated moved and multiplier, leading to various genome levels to scale-free insights.

Resolution Scale

The viewer has a large dynamic range in the size and resolution of the features it can display, from whole chromosomes/chromosome regions, to critical structure of critical importance for genome function. The viewer shows the visualization of 3D genomic structures and its project and link these to a classical linear representation.

Fig. 4: Dynamic zoom into the level of the DNA.

Fig. 5: Background image: Multi-chromosomal image (Ensembl) 15 to 21. Comparison, creation of entire new possibilities for understanding genome organization.

Fig. 6: Comparison of a simulated 3D-chromosomes/chromosome topology, combined with the as principle, linear information content in the DNA sequence and multi-dimensional mapping of Fig. 15.

Inter-Relations

In addition to the simultaneous mapping on one chromosome, the viewer allows the analysis of inter-chromosomal relationships based either on an external input (Fig. 2) or internal relation analysis (Fig. 3, 4, 6). Every genome region is a virtual 3D display, allowing the visualization of duplicon spreading and disease markers.

Features

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

Conclusion

The genome viewer presented here enables navigation to visualize and analyse the multi-dimensional aspects of genomes in a new dynamic way in combination with a non-linear chromosome view. A new GLOBE genome viewer, Prolonged, at the Erasmus Medical Center, an environment with active new inspiring research. It will be used extensively for future research, leading to a better understanding of the holistic properties of genomes, which is necessary for advanced diagnostic services and perhaps ultimate treatments.

Fig. 2: Multi-chromosomal relation view between duplication regions in and between chr. 15 & 21. Colored: duplication spreading degree

Fig. 3: Intra-chromosomal duplicon (Eichler et al.) compared to syndromes (blue/green), literature hot-spots (orange), and our classical approach (pink) in the chr. 22q.11 region

Fig. 5: Multi-chromosomal relation view between duplication regions in and between chr. 15 & 21. Colored: duplication spreading degree

Fig. 6: Correlation of a simulated 3D-chromosomes/chromosome topology, combined with the as principle, linear information content in the DNA sequence and multi-dimensional mapping of Fig. 15.
A New Holistic Genome Viewer for Molecular Genetics

Eussen, B., Moorhouse, M. J., Lesnussa, M., Muetgeert, M. & Knoch, T. A.


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 immense sequencing achievements and huge advances in the understanding of the general sequential, three-dimensional and regulatory organization. Here, we present the GLOBE 3D Genome Platform a completely novel grid based virtual “paper” tool and in fact the first system-biological genome browser integrating the holistic complexity of genomes in a single easy comprehensible platform: 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 and in a symbolic representation simultaneously and navigated by continuous scale-free zooming within a unique three-dimensional OpenGL and grid driven environment. In principle an unlimited number of 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. Any information can be searched and correlated by importing or calculating simple relations in real-time using grid resources. A general correlation and application platform for more complex correlative analysis and a front-end for system-biological simulations both using again the huge capabilities of grid infrastructures is currently under development. Hence, the GLOBE 3D Genome Platform is an example of a grid based approach towards a virtual desktop for genomic work combining the three fundamental distributed resources: i) visual data representation, ii) data access and management, and iii) data analysis and creation. Thus, the GLOBE 3D Genome Platform is the novel system-biology oriented information system urgently needed to access, present, annotate, and to simulate the holistic genome complexity in a unique gateway towards a real understanding, educative presentation and curative manipulation planning of this tremendous evolutionary information grail – genomes.

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 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 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, Brownian Dynamics, Monte Carlo, fluorescence in situ hybridization, confocal laser scanning microscopy, fluorescence correlation spectroscopy, super resolution microscopy, spatial precision distance microscopy, 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


