A New Holistic Genome Viewer for Molecular Cytogenetics

Bert Eussen [1,2], Michael J. Moorhouse [1,3], Michael Lesnussa [1-4], Maarten Muetgeert [1-4] and Tobias A. Knoch [1,4]

in cooperation with A. de Klein [2], P. v.d. Spek [3], F. G. Grosveld [5]

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

[2]Department of Clinical Genetics, [3]Department of Bioinformatics and [4]Biophysical Genomics Group in the [5]Department of Cell Biology
Erasmus Medical Center, Dr. Molewaterplein 50, NL- 3015 GE Rotterdam, The Netherlands
http://www.erasmusmc.nl/globe-consortium or globe@erasmusmc.nl

GLOBE - Consortium
Erasmus Medical Center Rotterdam
Towards Holistic Genomics

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 customizability 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.

Features

Flexible
Customizable
Intuitive Navigation

Real-Time Interaction & Analysis
Dynamical Resolution & Arrangement
Extremely Large & Multi-Dimensional Data

Bridge ALL Scales from Sequence to Morphology

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. in position, shape and colour. The viewer allows to visualize in prinziple an unlimited number of elements.

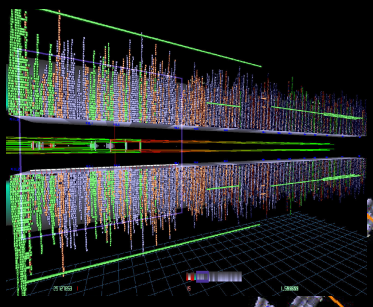


Fig. 1: Complete merged clone set (UCSC, NCBI, Ensembl) of chr. 15: colours represent association with duplicon regions.

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 correlation analysis (Fig. 1, 4, 6). Every genome dependent item is relatable e.g. syndromes to duplicons or genes families to breakpoints etc.

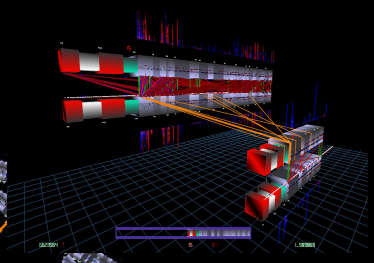
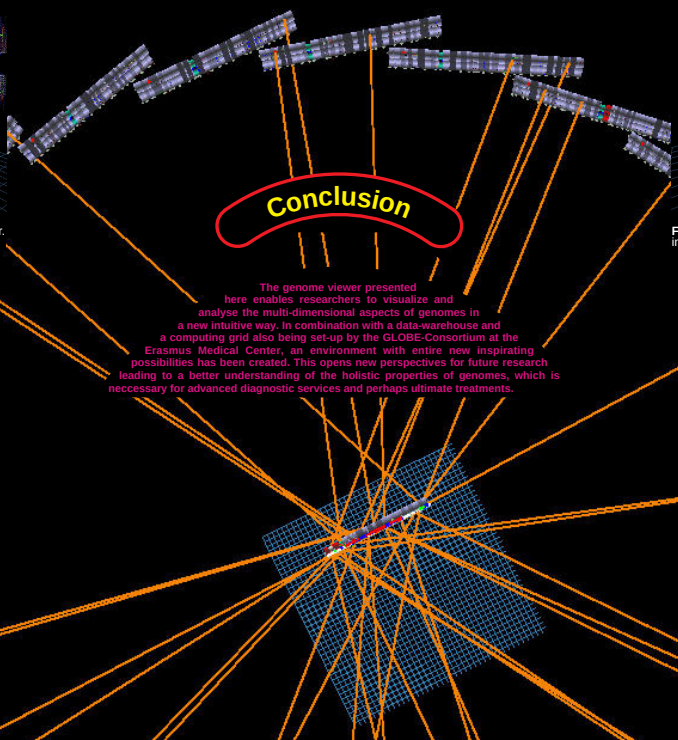


Fig. 2: Multi-chromosomal relation view between duplicon regions in and between chr. 15 & 21. Colours: duplicon spreading degree.

Conclusion

The genome viewer presented here enables researchers to visualize and analyse the multi-dimensional aspects of genomes in a new intuitive way. In combination with a data-warehouse and a computing grid also being set-up by the GLOBE-Consortium at the Erasmus Medical Center, an environment with entire new inspirating possibilities has been created. This opens new perspectives for future research leading to a better understanding of the holistic properties of genomes, which is neccessary for advanced diagnostic services and perhaps ultimate treatments.



Data Tracks

Syndrome	Chromosome
Break Points	Ideogram Bands
Duplicon	Chromatin Loops
Repeat Regions	Chromatin Fiber
Epigenetics	Histone
Genes /SNP	DNA

Data Tracks

BACS	3D-FISH
Fosmids	M-FISH
Genomic Arrays	CGH
Proteomic Arrays	Expression
Restriction Sites	3C
Primers	QPCR

Intra-Relations

Using the dynamic scaling range of the intra-chromosomal relationships can be studied in detail in relation to the track mapping (Fig. 1 & 2) concerning basic research, diagnostics and treatments. Assays can be projected, related, reviewed and redefined thus leading on various genome levels to scale-free insights.

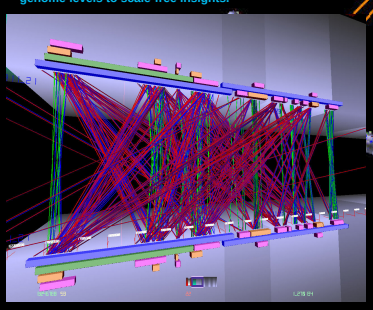


Fig. 3: Intra-chromosomal duplicons (Eichler et al.) compared to syndromes (blue/green), literature hot-spots (orange), and our defined hot-spots (pink) of the chr. 22q.11 region.

Resolution Scale

The viewer has a large dynamic range in the size and resolution of the features it can display: from whole chromosomes to individual bases. This new environment creates entire new possibilities for understanding genome organization.

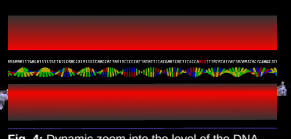


Fig. 4: Dynamic zoom into the level of the DNA.

Fig. 5: Background image: Multi-chromosomal relation between the breakpoints of chr. 15 to all other chromosomes. Colours: as in Fig. 2.

Structure

There are several physical levels of genetic information storage, e.g. DNA, chromatin and chromosomes. The interaction between information and the structural carrier is of critical importance for genome function. The viewer allows the visualisation of 3D genomic structures and to project and link these to a classical linear representation.

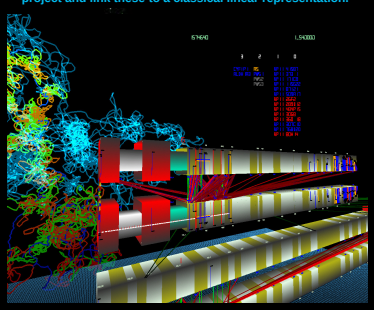


Fig. 6: Correlation of a simulated 3D chromatin/chromosome topology combined with the - in principle - linear information content in the DNA sequence and multi-dimensional mapping of chr. 15.

A New Holistic Genome Viewer for Molecular Genetics

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

High Resolution Biomedical Imaging - Symposium on the Introduction of the 4Pi Confocal Microscope, Erasmus Medical Center, Rotterdam, The Netherlands, 6th April, 2006.

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

Knoch, T. A. Dreidimensionale Organisation von Chromosomen-Domänen in Simulation und Experiment. (Three-dimensional organization of chromosome domains in simulation and experiment.) *Diploma Thesis*, Faculty for Physics and Astronomy, Ruperto-Carola University, Heidelberg, Germany, 1998, and TAK Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-010685-5 and ISBN 978-3-00-010685-9 (soft cover, 2rd ed.), ISBN 3-00-035857-9 and ISBN 978-3-00-0358857-0 (hard cover, 2rd ed.), ISBN 3-00-035858-7, and ISBN 978-3-00-035858-6 (DVD, 2rd ed.), 1998.

Knoch, T. A., Münkel, C. & Langowski, J. Three-dimensional organization of chromosome territories and the human cell nucleus - about the structure of a self replicating nano fabrication site. *Foresight Institute - Article Archive*, Foresight Institute, Palo Alto, *CA, USA,* http://www. foresight.org, 1- 6, 1998.

Knoch, T. A., Münkel, C. & Langowski, J. Three-Dimensional Organization of Chromosome Territories and the Human Interphase Nucleus. *High Performance Scientific Supercomputing*, editor Wilfried Juling, Scientific Supercomputing Center (SSC) Karlsruhe, University of Karlsruhe (TH), 27- 29, 1999.

Knoch, T. A., Münkel, C. & Langowski, J. Three-dimensional organization of chromosome territories in the human interphase nucleus. *High Performance Computing in Science and Engineering 1999*, editors Krause, E. & Jäger, W., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3-540-66504-8, 229-238, 2000.

Bestvater, F., **Knoch, T. A.**, Langowski, J. & Spiess, E. GFP-Walking: Artificial construct conversions caused by simultaneous cotransfection. *BioTechniques* 32(4), 844-854, 2002.

Knoch, T. A. (editor), Backes, M., Baumgärtner, V., Eysel, G., Fehrenbach, H., Göker, M., Hampl, J., Hampl, U., Hartmann, D., Hitzelberger, H., Nambena, J., Rehberg, U., Schmidt, S., Weber, A., & Weidemann, T. Humanökologische Perspectiven Wechsel - Festschrift zu Ehren des 70. Geburtstags von Prof. Dr. Kurt Egger. Human Ecology Working Group, Ruperto-Carola University of Heidelberg, Heidelberg, Germany, 2002.

Knoch, T. A. Approaching the three-dimensional organization of the human genome: structural-, scaling- and dynamic properties in the simulation of interphase chromosomes and cell nuclei, long- range correlations in complete genomes, *in vivo* quantification of the chromatin distribution, construct conversions in simultaneous co-transfections. *Dissertation*, Ruperto-Carola University, Heidelberg, Germany, and TAK†Press, Tobias A. Knoch, Mannheim, Germany, ISBN 3-00-009959-X and ISBN 978-3-00-009959-5 (soft cover, 3rd ed.), ISBN 3-00-009960-3 and ISBN 978-3-00-009960-1 (hard cover, 3rd ed.), ISBN 3-00-035856-9 and ISBN 978-3-00-010685-9 (DVD, 3rd ed.) 2002.

Knoch, T. A. Towards a holistic understanding of the human genome by determination and integration of its sequential and three-dimensional organization. *High Performance Computing in Science and Engineering 2003*, editors Krause, E., Jäger, W. & Resch, M., High-Performance Computing Center (HLRS) Stuttgart, University of Stuttgart, Springer Berlin-Heidelberg-New York, ISBN 3- 540-40850-9, 421-440, 2003.

Wachsmuth, M., Weidemann, T., Müller, G., Urs W. Hoffmann-Rohrer, **Knoch, T. A.**, Waldeck, W. & Langowski, J. Analyzing intracellular binding and diffusion with continuous fluorescence photobleaching. *Biophys. J.* 84(5), 3353-3363, 2003.

Weidemann, T., Wachsmuth, M., **Knoch, T. A.**, Müller, G., Waldeck, W. & Langowski, J. Counting nucleosomes in living cells with a combination of fluorescence correlation spectroscopy and confocal imaging. *J. Mol. Biol.* 334(2), 229-240, 2003.

Fejes Tóth, K., **Knoch**, **T. A.**, Wachsmuth, M., Frank-Stöhr, M., Stöhr, M., Bacher, C. P., Müller, G. & Rippe, K. Trichostatin A induced histone acetylation causes decondensation of interphase chromatin. *J. Cell Science* 177, 4277-4287, 2004.

Ermler, S., Krunic, D., **Knoch, T. A.**, Moshir, S., Mai, S., Greulich-Bode, K. M. & Boukamp, P. Cell cycle-dependent 3D distribution of telomeres and telomere repeat-binding factor 2 (TRF2) in HaCaT and HaCaT-myc cells. *Europ. J. Cell Biol.* 83(11-12), 681-690, 2004.

Kost, C., Gama de Oliveira, E., **Knoch, T. A.** & Wirth, R. Spatio-temporal permanence and plasticity of foraging trails in young and mature leaf-cutting ant colonies (*Atta spp.*). *J. Trop. Ecol.* 21(6), 677- 688, 2005.

Winnefeld, M., Grewenig, A., Schnölzer, M., Spring, H., **Knoch, T. A.**, Gan, E. C., Rommelaere, J. & Cziepluch, C. Human SGT interacts with BAG-6/Bat-3/Scythe and cells with reduced levels of either protein display persistence of few misaligned chromosomes and mitotic arrest. *Exp. Cell Res.* 312, 2500-2514, 2006.