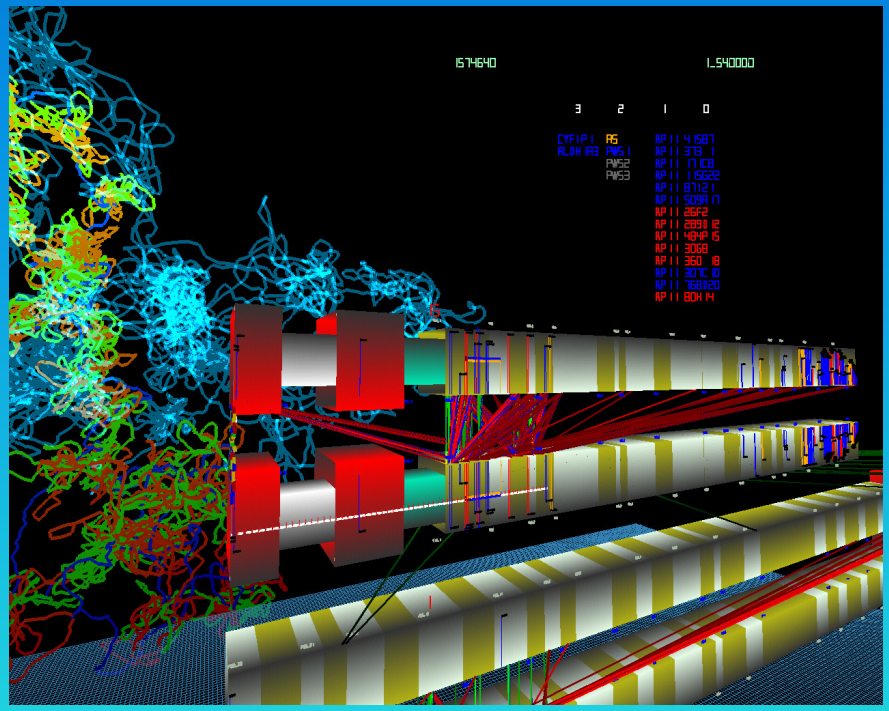
The GLOBE 3D Genome Browser

GLOBE Genome Information Systems Solutions

to navigate and discover the secrets of genomes



Ing. Hubert B. Eussen, Dr. Michael J. Moorhouse, & Dr. Tobias A. Knoch

Prof. Dr. A. de Klein, Prof. Dr. A. Uiterlinden, & Prof. Dr. F. G. Grosveld

Dept. Clinical Genetics, Dept. Internal Medicine, Dept. Cell Biology & Genetics

Erasmus Medical Center

The Adventure to Discover the Secrets of Life

Better Health and Longer Living

Genetic information is at the heart of every living being !

Genome Research, Products & Diagnostics

- extremely huge
- extremely complicated

>>> ***The Human Genome***

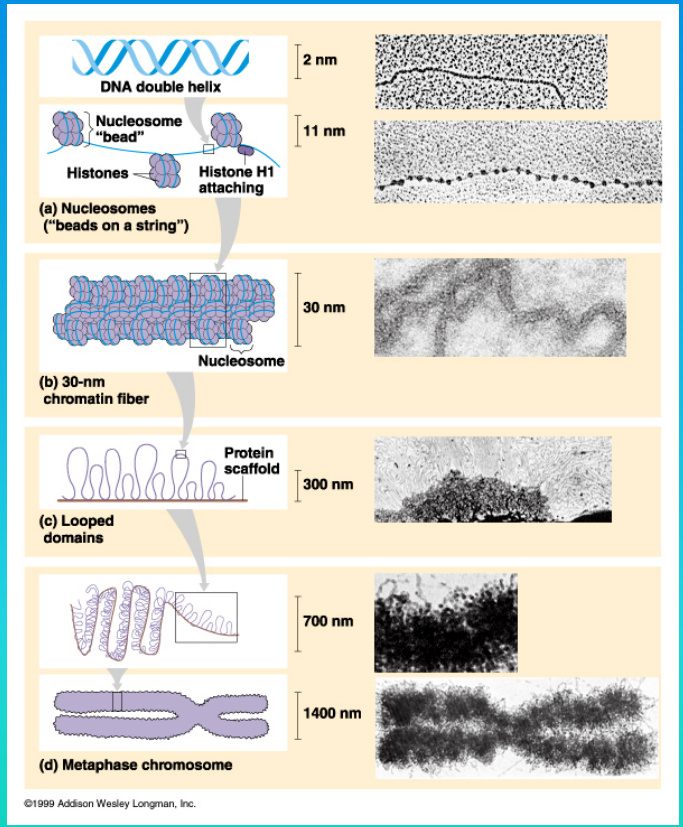
- 3,000,000,000 base pairs
- 30,000 genes
- >1,000,000 disease relevant mutations

Research & Development

- genome organization and function
- genotype-phenotype correlation
- personal genome projects

Genomic Products & Diagnostics

- >5,000 genetic tests on the market
- per test often >10,000 genomic markers



>>> e.g. breast cancer, cardiovascular disease, prenatal diagnostics

The Challenge to Tame the Genomic Beast

The Requirements/Needs of the Genomic Sector

Adequate Platform to Exploit the Genomic Era !!

Generals

- visualization of data and relations
- combining public and experimental data
- integration from base pair to chromosome

Research & Development

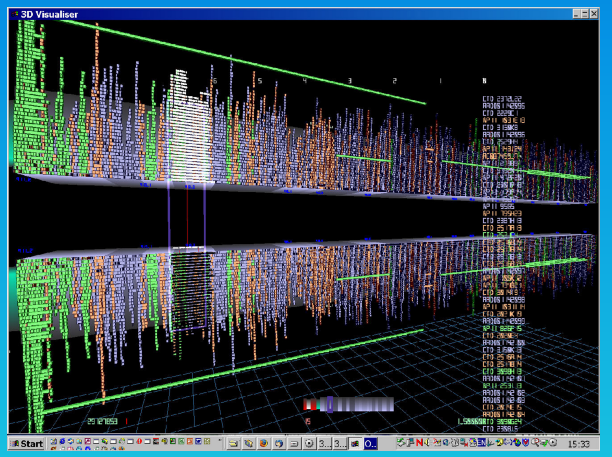
- working at the knowledge frontier
- flexibility, customizable, real-time
- annotation of data and experiment design

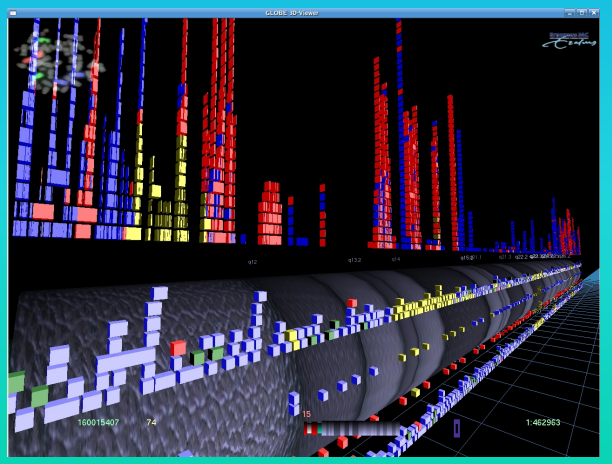
Genomic Product Suppliers

- catalog and sales of complex products
- customizable products on demand
- comparison to other products

Diagnostic Institutions

- easy visual access to diagnostic results
- links to disease data bases for diagnosis





Solution

Our New Standard for Integrated Scientific, Clinical and Industrial Genome Research, Application and Exploitation

New Concept of Data Navigation

- intuitive and easy to use
- flexible and generic
- multi-dimensional and real-time

Customer Benefits

- handier and faster to work with
- things which could not be done before
- entire new options for customers

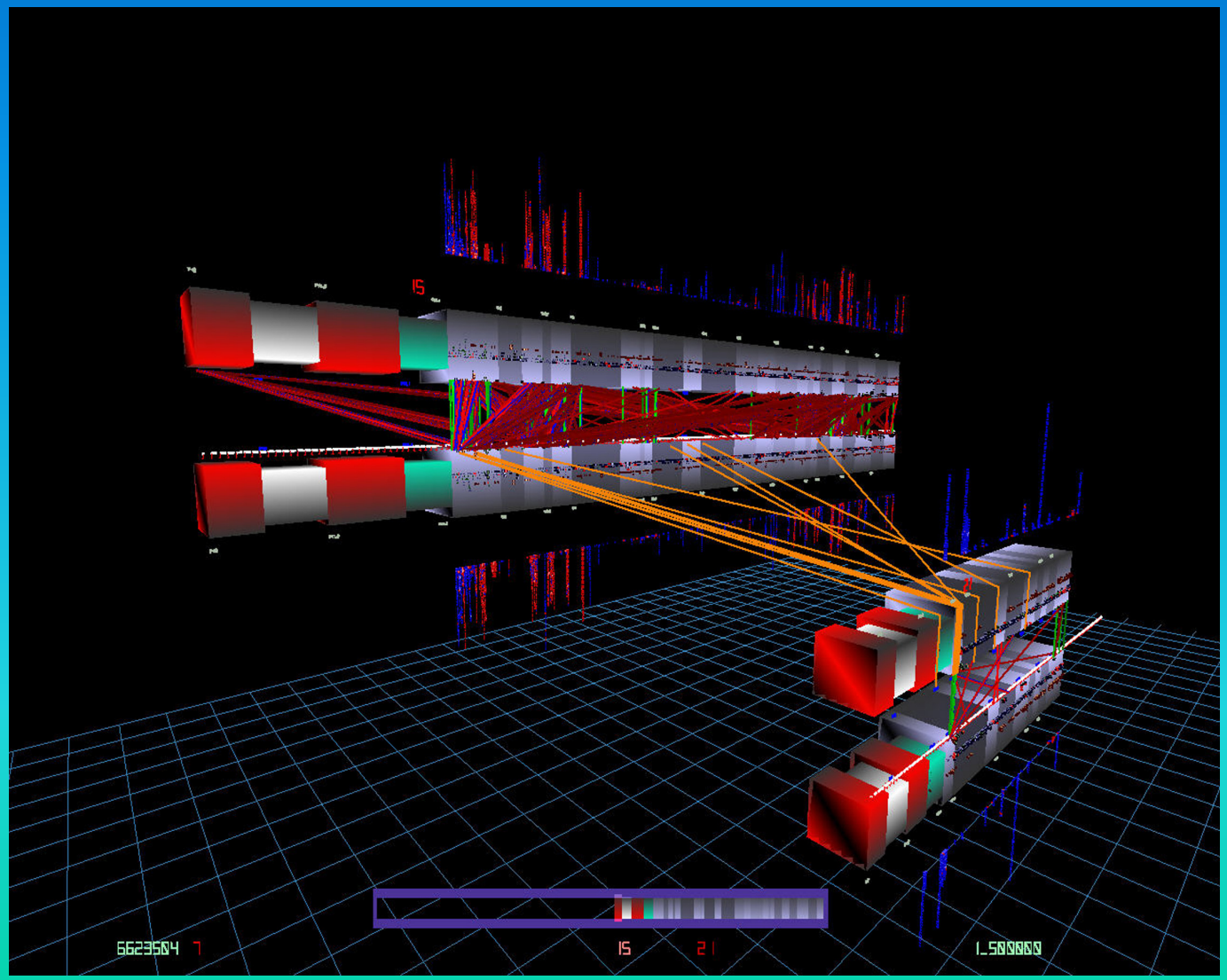
Rationale: A Genomic Global Positioning System (GPS)

>>> street list >>> ATLAS >>> Google Earth >>> electronic TOMTOM

The GLOBE 3D Genome Browser is the Genome Positioning System and Google Earth / Excel / Photoshop in a Single Platform!

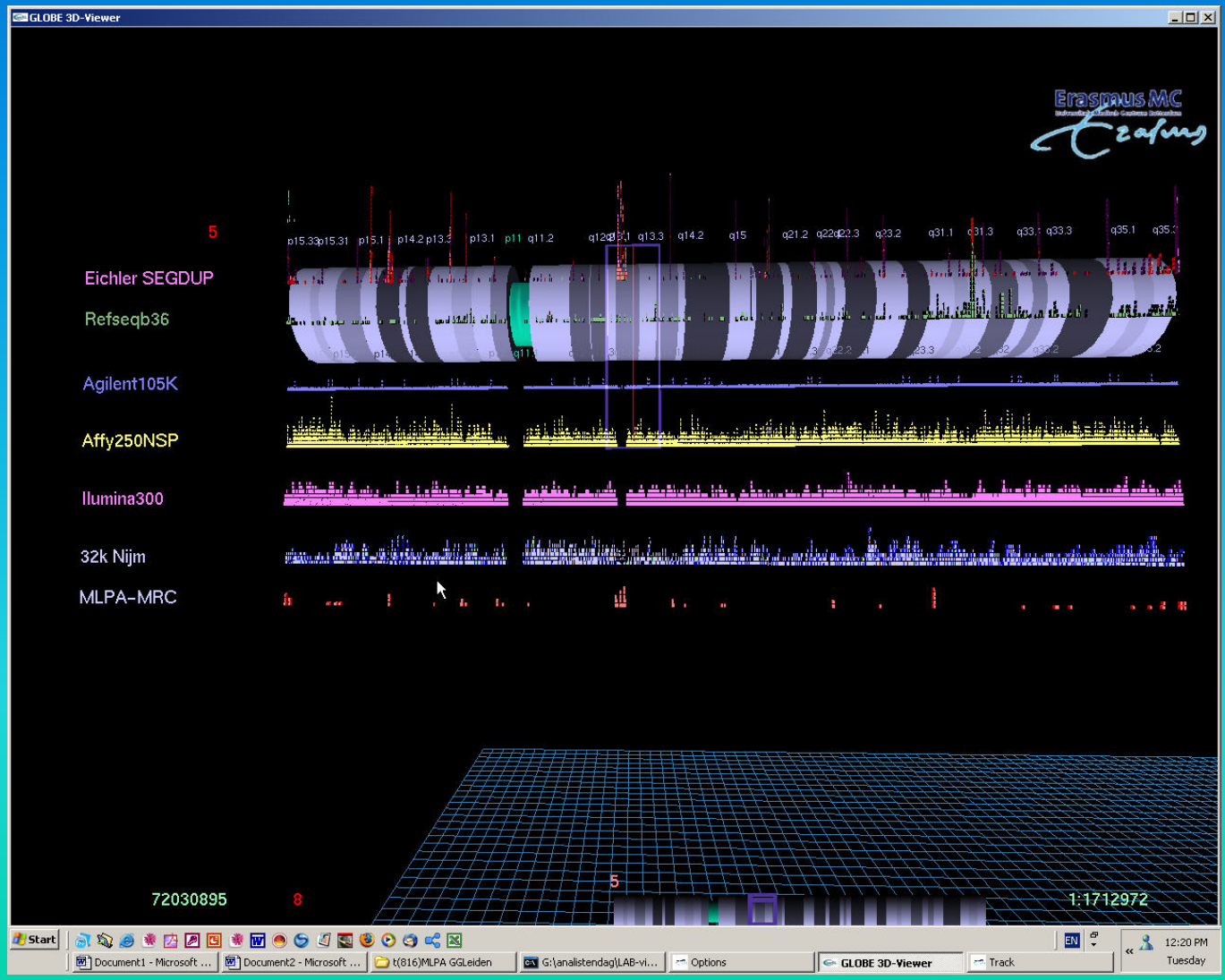
Relations to Look at, are the Most Sexy Ones

The Human Visual System is Still Better than any Program !



The Neighbors Wife is Always Attractive!

Comparison of Different Data/Experiments is the Key in Genomics !



The Devils Sitting on the Cake

Current Solutions in the Genomic Sector

Properties of Current Solutions

- difficult to use and long training
- non generic and home-brewn
- integration from base pair to chromosome

The Witches

- content organizations: Ensemble, UCSD
- home made solutions and small business
- non-classical competitors: Google, Adobe

Our Wisdom Team

- > 20 years of experience in genomics
- 3 expertiese covering 100% of genomics
- knowledge at the scientific and clinical frontier

>>> We outperform by:

- knowledge leadership
- technological leadership

The Cave of the Gold and Diamonds

The Market Opportunities in the Genomic Sector

"I want this tool: NOW"

Evin Eichler
Big Shot in Genomics
Washington

"Every CEO wants to have it on a USB"

ENZO
general product manager

"Your are 10 years ahead of the field"

Viviane Reding
EU Commissioner
Information Society & Media

General Market Size

- > 100,000,000 species
- > 7,000,000,000 individual
- > 1,000,000 disease relev

Research ... elopment Sector

- ~500,000 genomic ...kers
- >10,000 academic institu...
- >10,000 BioTech / Pharma R...

...rking in genomics
...centres
...Pharma R&D
...uros per year

- ~500,000 diagnostic appliers
- >100,000 diagnostic Labs
- >5,000 BioTech/Pharma Suppliers

In The Netherlands

- >2000 diagnostic workers
- >100 diagnostic labs
- >300 diagnostic companies present
- >100,000 clinical cases per year

>>> The Market has Huge Growth Potential > 50 years !
>>> The Genomic Sector will not loose its importance !

The Gold and The Diamonds

Financial Opportunities and Return of Investments

Market Opportunities

- licenses
- white label products
- content sale
- advertising
- e-commerce

Individual R&D licence

- academic: 1,000 Euro/a
- industry: 5,000 - 20,000 Euro/a

Corporate R&D licence

- academic: 50,000 - 200,000 Euro/a
- industry: 100,000 - 200,000 Euro/a

Genomic Product Suppliers

- sales portal: 100,000 - 200,000 Euro
- sales share: 2 - 5 % of turnover

Diagnostic Institutions

- individual licence: 3,000-7,000 Euro/a

We have already:

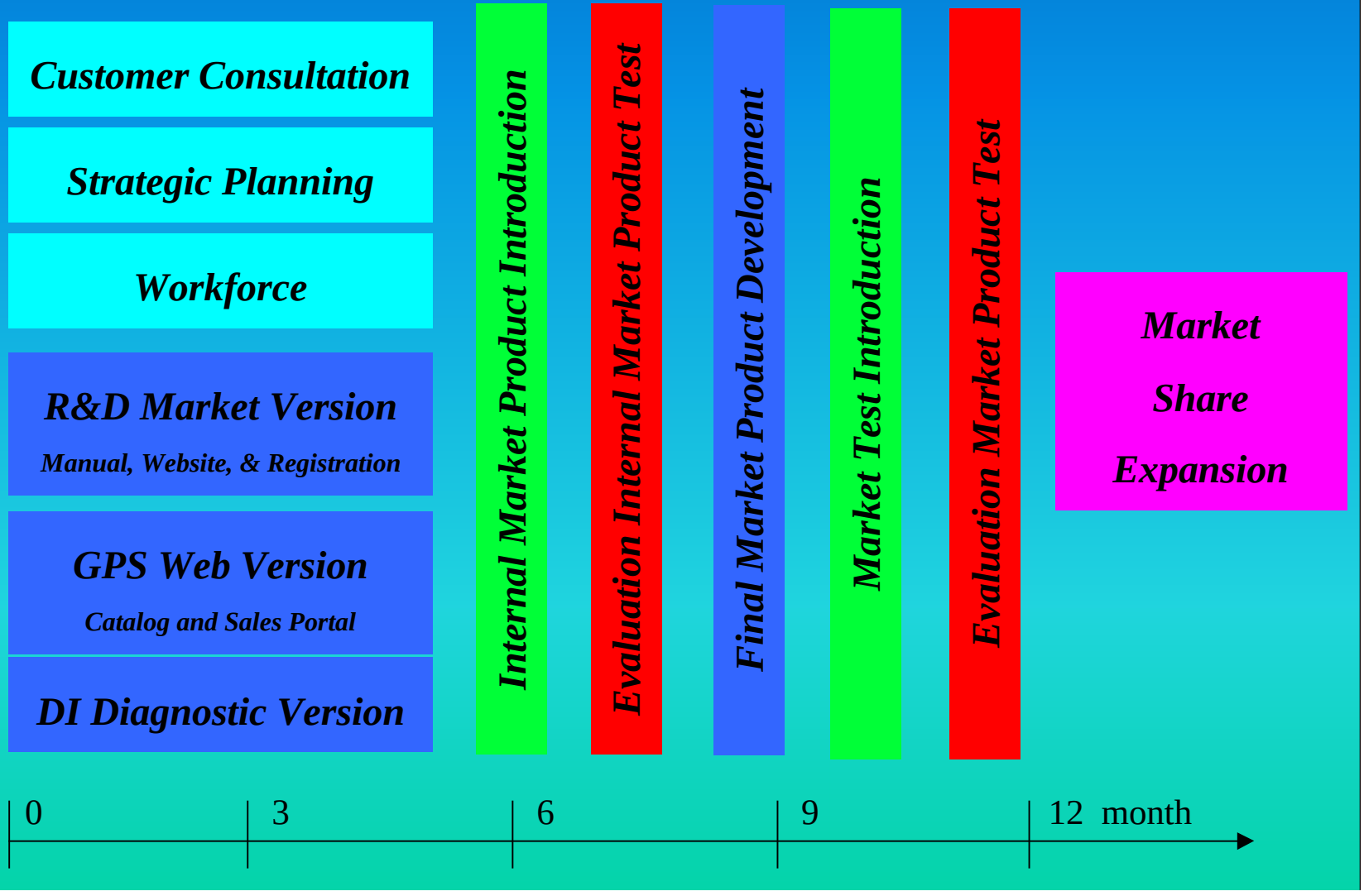
- >10 Erasmus MC internal customers (on a cooperation basis)
- >10 cooperation users world wide (on a cooperation basis)
- the German MediGRID as a corporation partner

- Erasmus MC Incubator management financing: 20,000 + 80,000 Euro

It is a SHORT Way to Tipperary

12 Month Plan for a (Test) Market Introduction

Budget: For the WHOLE Process Cheap 200 kEuro.



The GLOBE 3D Genome Browser

Knoch, T. A., Eussen, B. & Moorhouse, M.

NBIC Venture Challenge, University of Neijmegen, Neijmegen, The Netherlands,
19th June, 2007.

Abstract

The combination of genome sequence and structure, its annotation and experimental data in an accessible and comprehensible way is a major challenge. Increasingly, there are 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 huge limitations concerning their focus on linear presentation, standardized input and data bank accessibility. Also customisability by a remote user with special requirements is difficult.
Therefore, the GLOBE-Consortium has developed the next generation genome viewer - the GLOBE 3D-Viewer - to visualize multi-dimensional data sets from various sources in an easy intuitive manor while accounting for the complexity of genome organization. Together with an easily accessible data-warehouse for archiving (un-)published experimental data and the Erasmus Computing Grid (ECG) for fast interactive large-scale (correlative) analysis the GLOBE-Consortium creates an intuitive holistic environment for genome research. Thus, the GLOBE 3D-Viewer sets the new standard to integrate complex genomic data sets within a single holistic display system and thus is major advance for scientific and clinical genome research.

Rational behind the GLOBE 3D-Viewer

The GLOBE 3D-Viewer is based on many a basic rational driven by the complex nature of genome organization in general and by the diverse individual research and diagnostic data types in particular. Thus, the key features of the GLOBE 3D-Viewer are:

- flexibility, customisability, intuitive navigation
- real-time interaction and analysis
- dynamic scaling, semantic zooming, and object arrangement
- visualization of extremely large data sets and multi-dimensional data
- bridging all scales of genome organization from sequence to morphology

Capabilities of the Prototype GLOBE 3D-Viewer

The GLOBE 3D-Viewer allows the integration of all possible linear and structural genomic types of information from the DNA sequence to complete genomes in a three-dimensional environment. Genomic elements are visualized in a flexible and customisable manor and can be arranged in relation to each other as desired. The navigation within this three-dimensional space is intuitive and resembles the virtual environments of flight simulators or architectural planning tools. This supports also rapid resolution change, i.e. zooming-out from the base pair to the whole-genome level is possible within an eye-glimpse using fly-in/out paradigm while including

appearance or relational positioning changes of genetic elements. The GLOBE 3D-Viewer is capable of relating extremely large numbers of genomic elements and multi-dimensional data-sets to each other with inbuilt analysis running in real-time.
In addition, the GLOBE 3D-Viewer is able to present the three-dimensional organisation of genomes from base pair resolution to complete chromosomes. This allows the direct projection of linear genomic information on its real three-dimensional spatial architecture for the first time.

The GLOBE 3D-Viewer integrates e.g. the following genomic and experimental elements or data types in a three-dimensional virtual environment:

- *Genomic information types:* Syndromes, break-points, duplicons, repeat regions, epigenetic annotations, genes and single nucleotide polymorphisms.
- *Structural genome features:* Chromosomes, ideogram bands, chromatin loops, chromatin fiber conformation, nucleosomes, DNA double helix.
- *Experimental genome features:* Bacterial Artificial Chromosomes, Yeast Artificial Chromosomes, Fosmids, genomic and proteomic arrays, restricition sites, oligomers, primers.
- *Genomic experiment types:* 3D fluorescence *in-situ* hybridization (3D-FISH), Metaphase fluorescence *in-situ* hybridization (M-FISH), comparative genome hybridisation (CGH), expression profiling, quantitative polymerase chain reaction (QPCR).

The GLOBE 3D-Viewer creates entire new possibilities in the visualization in the holistic properties of genomes. It opens new perspectives for future research leading to a better understanding of genomes and their system biological aspects. This is of major importance in respect to advanced diagnostics and ultimate disease treatments.

Opportunities of the GLOBE 3D-Viewer

The GLOBE 3D-Viewer creates a new intuitive virtual environment to show for the first the enormous complexity of genomes in a *single display*. This leads to an all encompassing understanding of genomes and opens exciting new opportunities in every aspect of scientific, clinical, education as well as commercial applications:

- *Scientists* benefit from easier planning and experimental analysis as well as hypothesis building and better comprehension of genomic function.
- *Diagnostic users* benefit from a better overview of their diagnostic tasks and the evaluation of the results including patient counselling and treatment.
- *Educators and publishers* benefit from enhanced presentation of genome organisation and function in classrooms or e-publications.
- *Commercial applicants* benefit from improved planning of new products, their in test, increased sales as well as more appealing marketing of their products.

In summary, the GLOBE 3D-Viewer offers the opportunity to understand genomes on the necessary system-biological level for the first time, and thus sets the new standard for genome browsers for scientific and clinical research and as well as for unseen possibilities concerning educational and commercial purposes. Therefore, the GLOBE 3D-Viewer creates also unrivalled chances for world-wide introduction generating a sustainable revenue stream.

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.

Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Yassene, M., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. Grid-basierte Services für die elektronische Patientenakte der Zukunft. *E- HEALTH-COM - Magazin für Gesundheitstelematik und Telemedizin*, 4(2), 61-63, 2007.

de Zeeuw, L. V., **Knoch, T. A.,** van den Berg, J. & Grosveld, F. G. Erasmus Computing Grid - Het bouwen van een 20 TeraFLOP virtuelle supercomputer. *NIOC proceedings 2007 - het perspective of lange termijn*. editor Frederik, H. NIOC, Amsterdam, The Netherlands, 52-59, 2007.

Rauch, J., **Knoch, T. A.** , Solovei, I., Teller, K. Stein, S., Buiting, K., Horsthemke, B., Langowski, J., Cremer, T., Hausmann, M. & Cremer, C. Lightoptical precision measurements of the Prader- Willi/Angelman Syndrome imprinting locus in human cell nuclei indicate maximum condensation changes in the few hundred nanometer range. *Differentiation* 76(1), 66-82, 2008.

Sax, U., Weisbecker, A., Falkner, J., Viezens, F., Mohammed, Y., Hartung, M., Bart, J., Krefting, D., **Knoch, T. A.** & Semler, S. C. Auf dem Weg zur individualisierten Medizin - Grid-basierte Services für die EPA der Zukunft. *Telemedizinführer Deutschland 2008*, editor Jäckel, A. Deutsches Medizinforum, Minerva KG, Darmstadt, ISBN 3-937948-06-6, ISBN-13 9783937948065, 47-51, 2008.

Drägestein, K. A., van Capellen, W. A., van Haren, J. Tsibidis, G. D., Akhmanova, A., **Knoch, T. A.**, Grosveld, F. G. & Galjart, N. Dynamic behavior of GFP-CLIP-170 reveals fast protein turnover on microtubule plus ends. *J. Cell Biol. 180(4)*, 729-737, 2008.

Jhunjhunwala, S., van Zelm, M. C., Peak, M. M., Cutchin, S., Riblet, R., van Dongen, J. J. M., Grosveld, F. G., **Knoch, T. A.**$^{+}$ & Murre, C.$^{+}$ The 3D-structure of the Immunoglobulin Heavy Chain Locus: implications for long-range genomic interactions. *Cell 133(2)*, 265-279, 2008.

Krefting, D., Bart, J., Beronov, K., Dzhimova, O., Falkner, J., Hartung, M., Hoheisel, A., **Knoch, T. A.**, Lingner, T., Mohammed, Y., Peter, K., Rahm, E., Sax, U., Sommerfeld, D., Steinke, T., Tolxdorff, T., Vossberg, M., Viezens, F. & Weisbecker, A. MediGRID - Towards a user friendly secured grid infrastructure. *Future Generation Computer Systems 25(3)*, 326-336, 2008.