Erasmus Computing Grid

From Sequence to Morphology

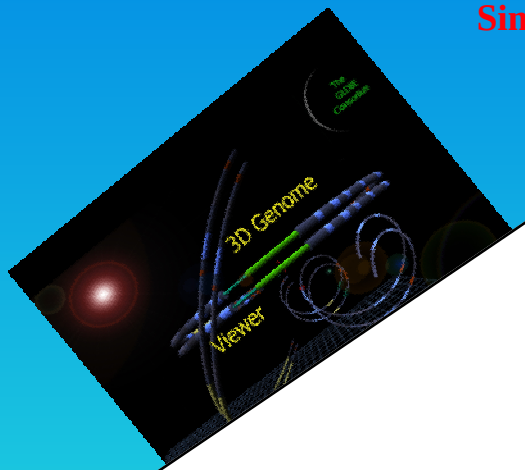
-

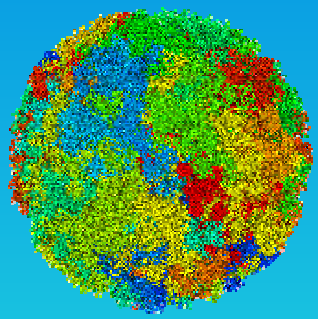
Approaching the Three-Dimensional Org… Genome

Simulation of Structu… …es

Th…

…wer





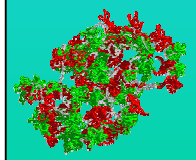
Tobias A. Knoch

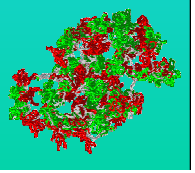
…ysical Genomics in Cell Biology & Genetics

B… …ael Moorhouse, Luc de Zeeuw, Annelies de Klein, & Frank Grosveld

Cell Biology & Genetics, Clinical Genetics, & Virology

Erasmus MC, Rotterdam,The Netherlands





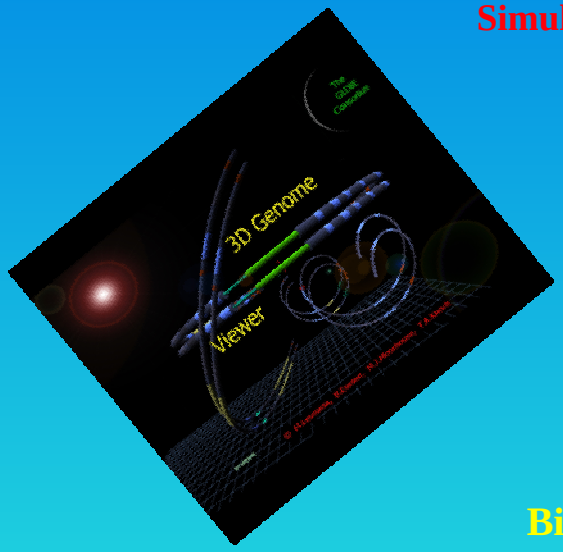
Erasmus Computing Grid

From Sequence to Morphology

Erasmus MC
University Medical Center Rotterdam

-

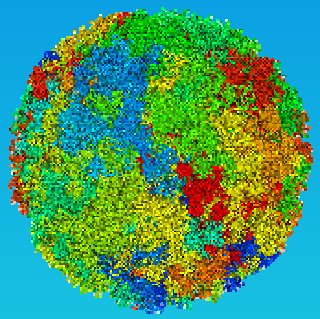
Approaching the Three-Dimensional Organization of the Human Genome



Simulation of Structural-, Scaling and Dynamic-Properties

The Erasmus Computing Grid

The Next Generation Genome Viewer



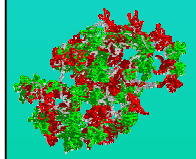
by

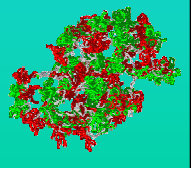
Tobias A. Knoch

Biophysical Genomics in Cell Biology & Genetics

Bert Eussen, Michael Moorhouse, Luc de Zeeuw, Annelies de Klein, & Frank Grosveld

Cell Biology & Genetics, Clinical Genetics, & Virology





Erasmus MC, Rotterdam,The Netherlands

Erasmus Computing Grid

Erasmus MC

HOGESCHOOL ROTTERDAM

From Sequence to Morphology

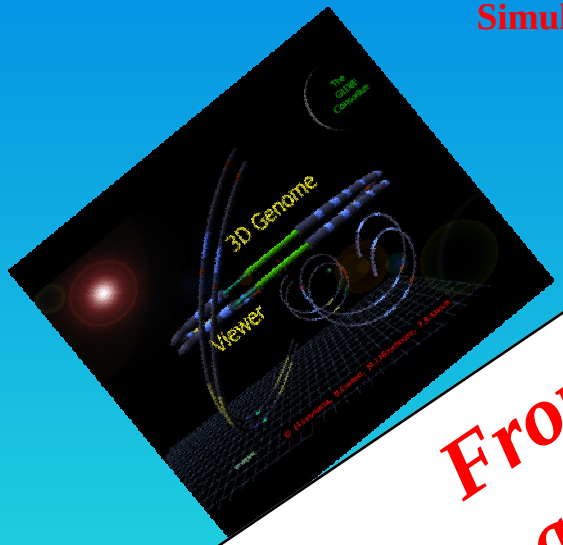
-

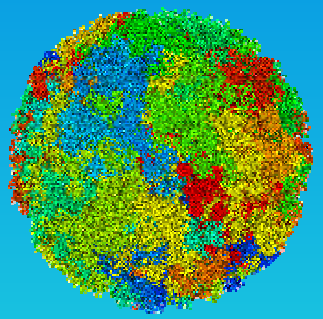
Approaching the Three-Dimensional Organizati… Genome

Simulation of Structural-, Scalin…

The Erasmu…

The Next G…





From Sequence to Morphology: Towards a Holistic Understanding of Genomes!

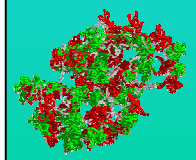
…bias A. Knoch

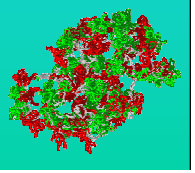
…cal Genomics in Cell Biology & Genetics

B… Moorhouse, Luc de Zeeuw, Annelies de Klein, & Frank Grosveld

Cell Biology & Genetics, Clinical Genetics, & Virology

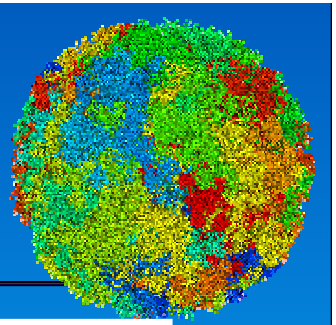
Erasmus MC, Rotterdam,The Netherlands

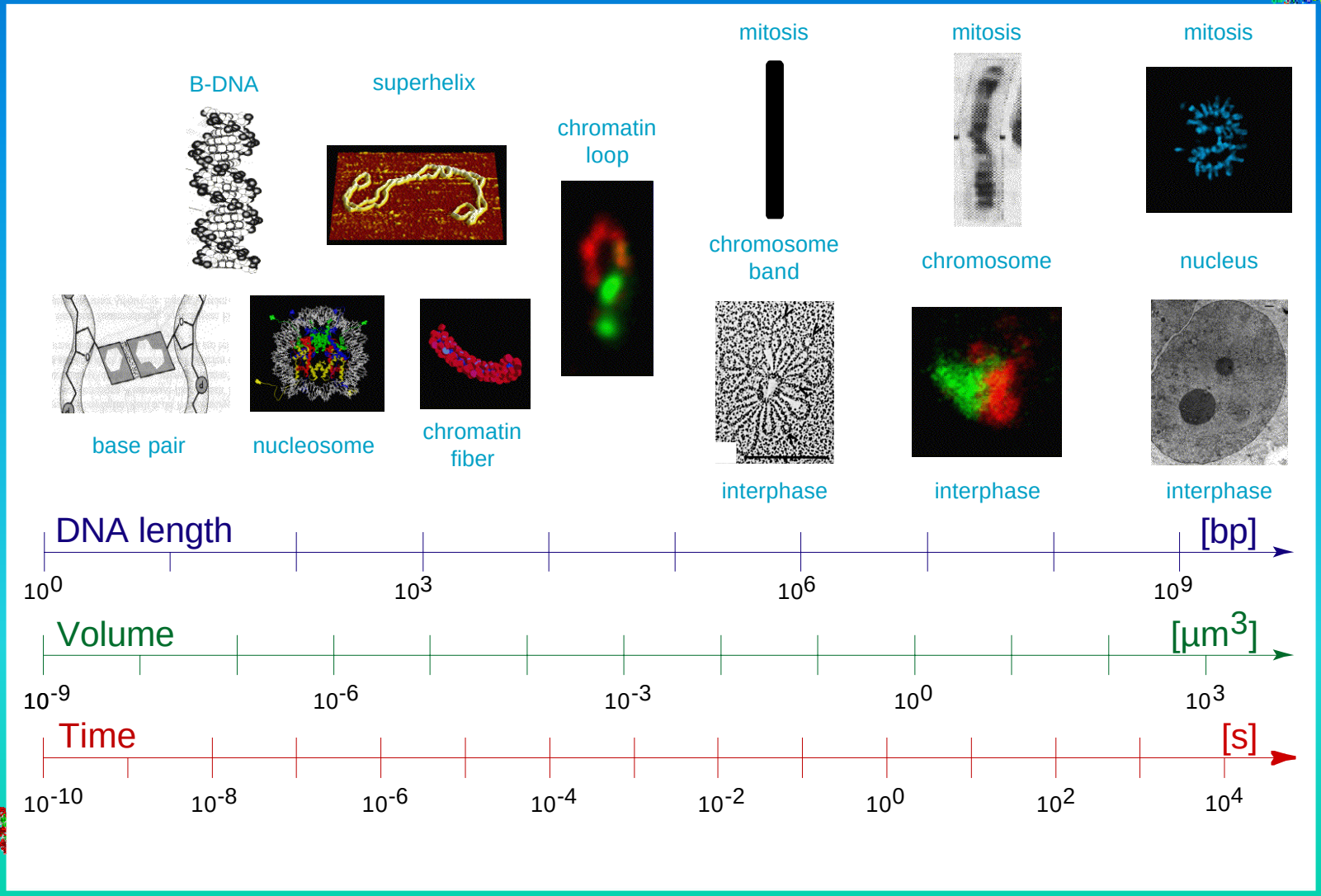


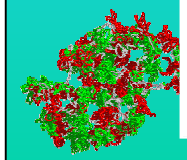


Dynamic and Hierarchical Genome Organization

The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogeneses, or cancer development is still unclear?

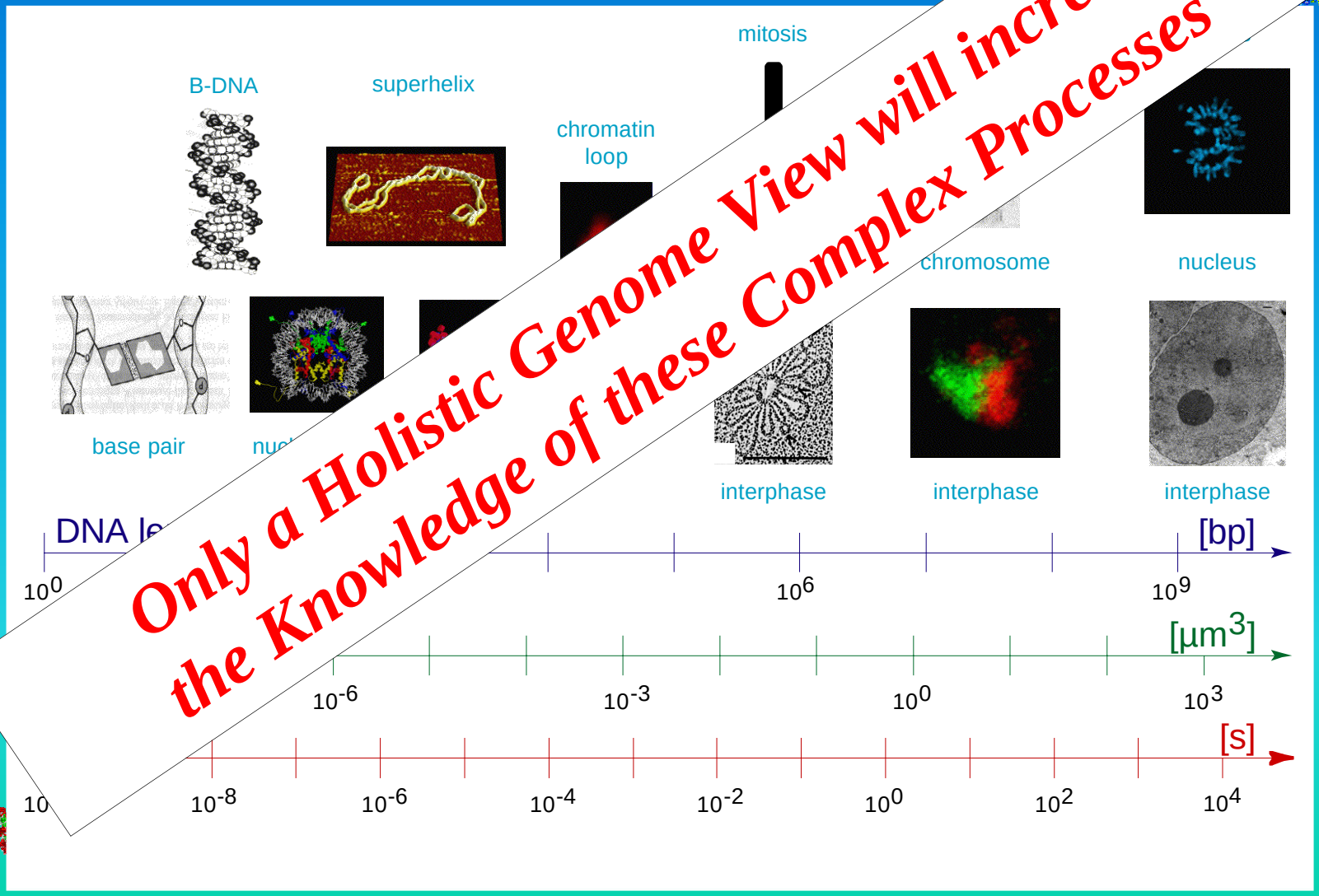






Dynamic and Hierarchical Genome Organization

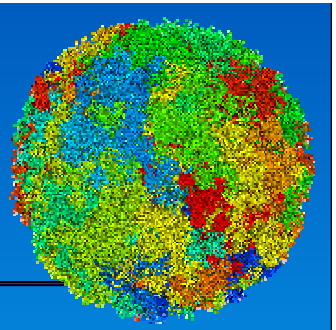
The different organization levels of genomes bridge several orders of magnitude concerning space and time. How all of these organization levels connect to processes like gene regulation, replication, embryogeneses, or cancer development is still unclear?



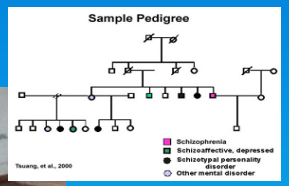
Only a Holistic Genome View will increase the Knowledge of these Complex Processes

The Complexity of Cytogenetic Diagnostics

The process of cytogenetic analysis requires proper patient and sample analysis as well asa comprehensive evaluation of the results.

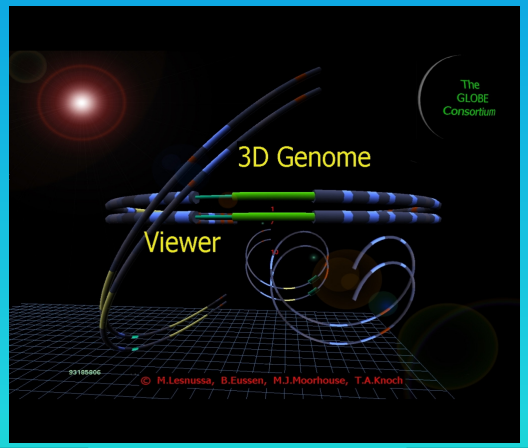


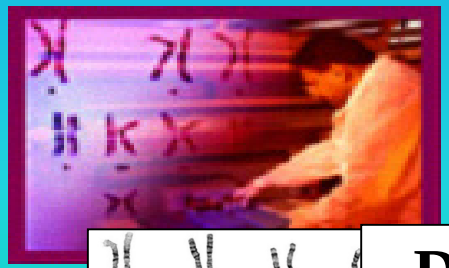
Patient

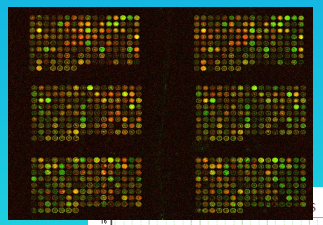


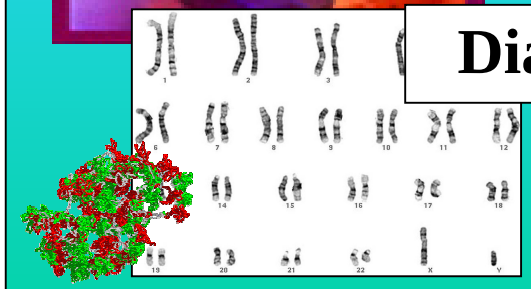
Treatment

Sample



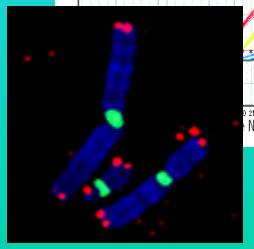


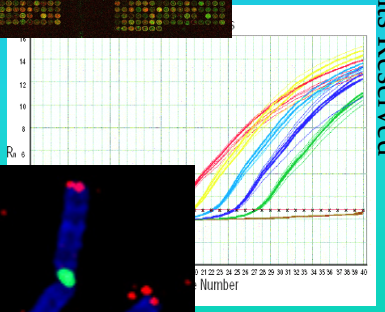




Diagnosis

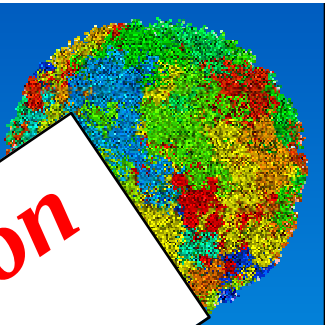
Analysis



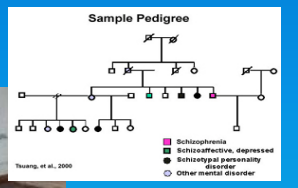


The Complexity of Cytogenetic Diagnostics

The process of cytogenetic analysis requires proper patient and sample analysis as well asa comprehensive evaluation of the results.

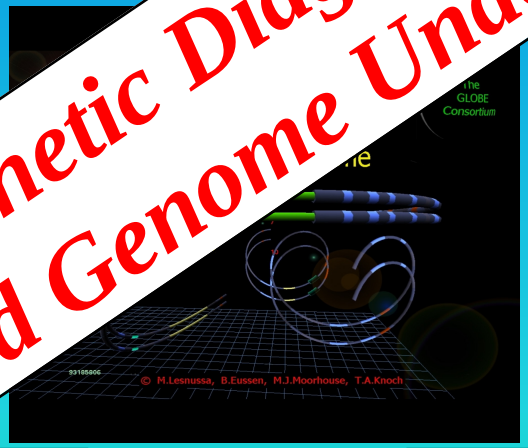


Patient

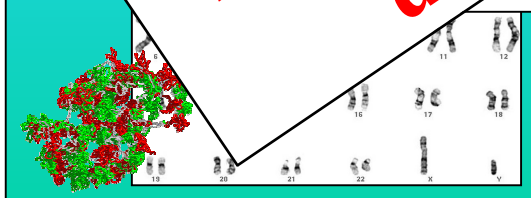


Treatment

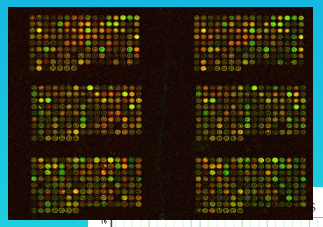
Sample

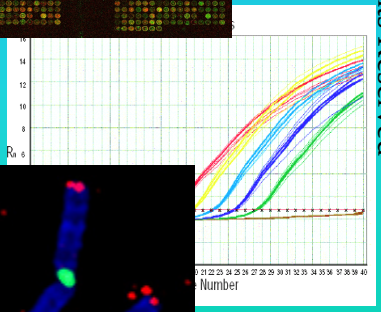


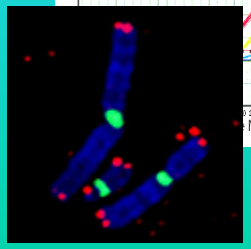
Diagnosis



Analysis





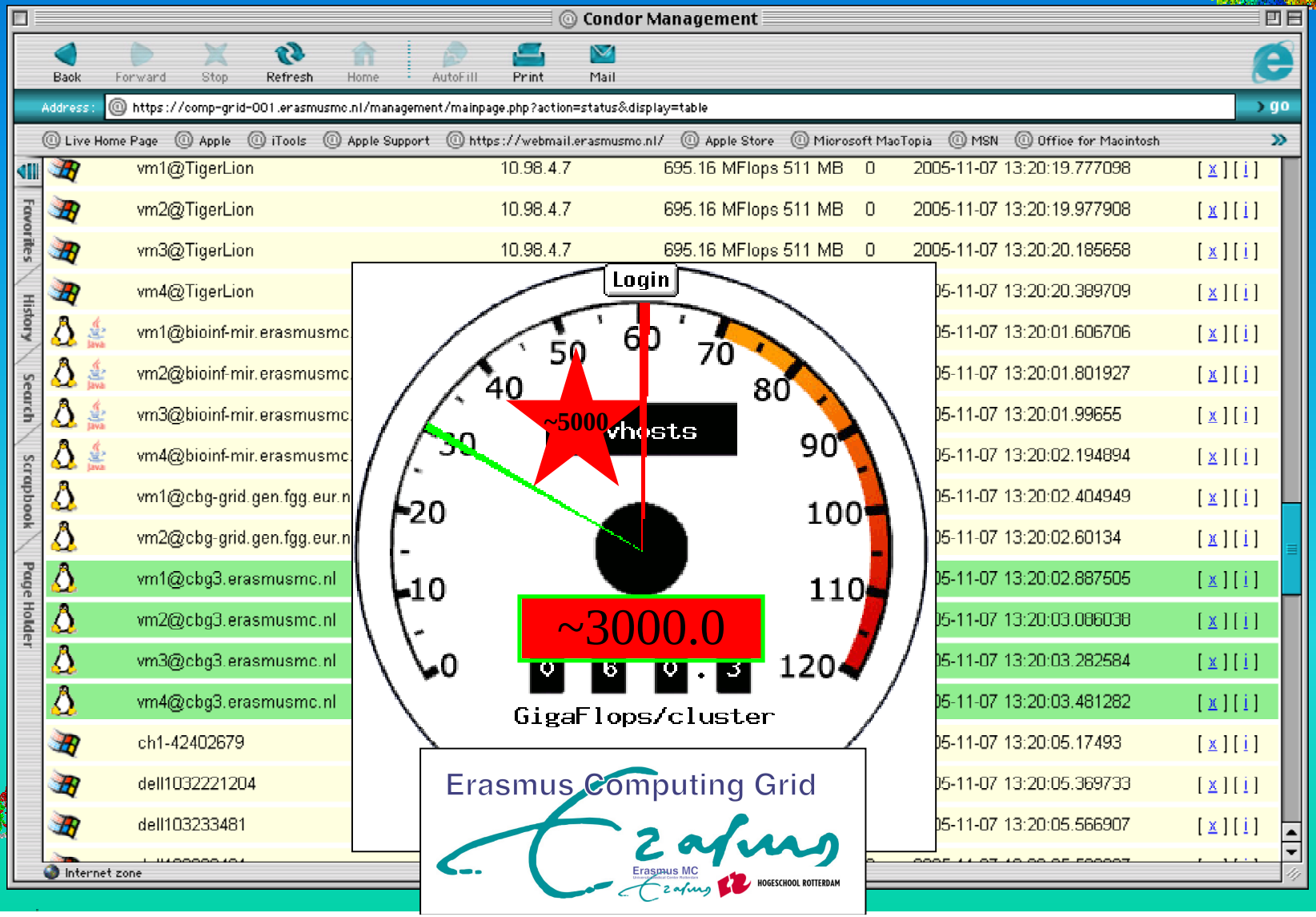


Better Cytogenetic Diagnostics is based on advanced Genome Understanding!

The Erasmus Computing Grid

Building a Super-Computer at Erasmus MC for Free!

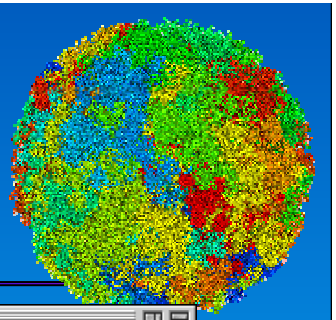
Applications at Erasmus MC: E.g. genome sequence analysis, protein structure simulations, chip/array analysis, epidemiology of viral infections, patient diagnostic image analysis (AMI).

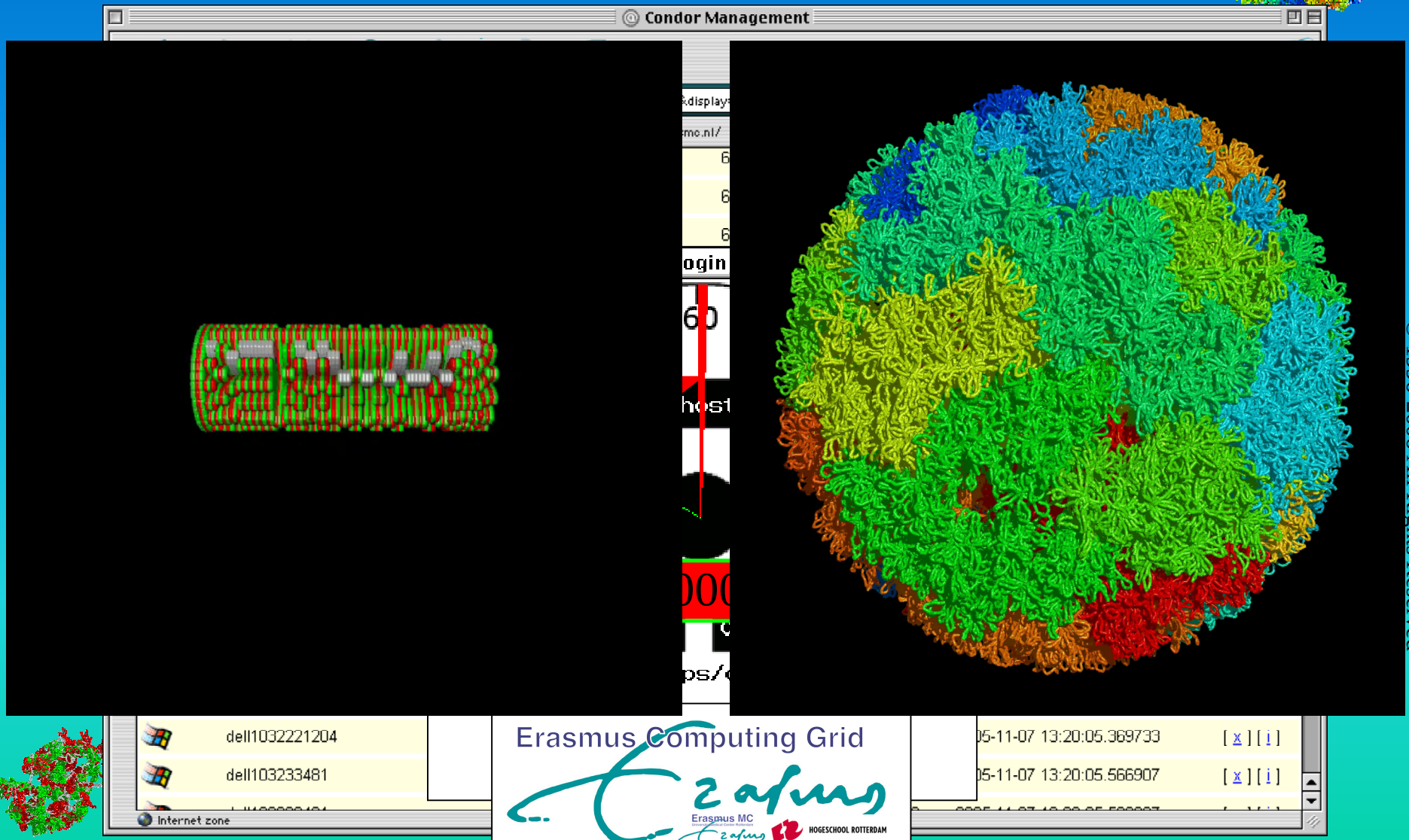


The Erasmus Computing Grid

Building a Super-Computer at Erasmus MC for Free!

Applications at Erasmus MC: E.g. genome sequence analysis, protein structure simulations, chip/array analysis, epidemiology of viral infections, patient diagnostic image analysis (AMI).





Erasmus Computing Grid

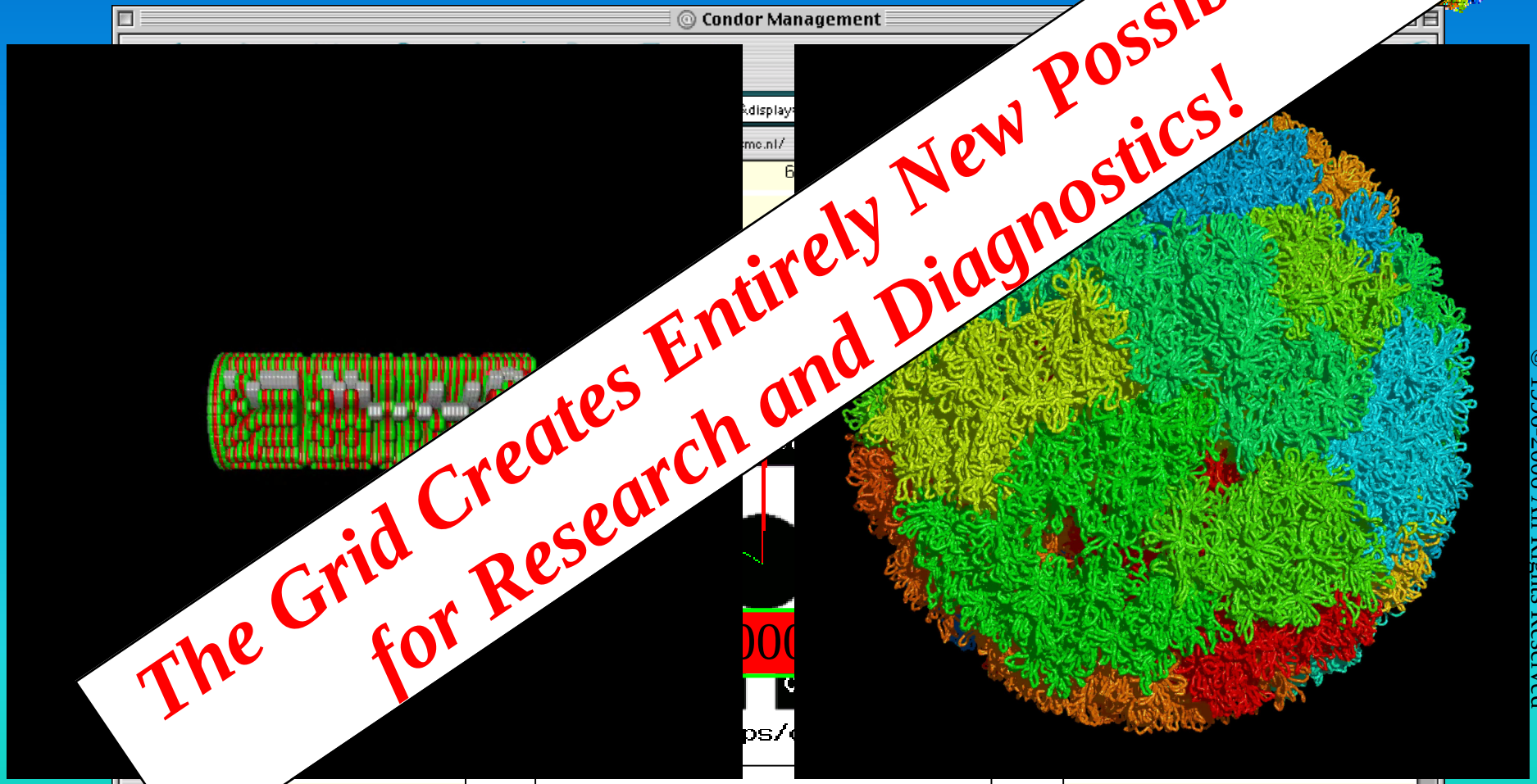
Erasmus MC

HOGESCHOOL ROTTERDAM

The Erasmus Computing Grid

Building a Super-Computer at Erasmus MC for Free!

Applications at Erasmus MC: E.g. genome sequence analysis, protein structure simulations, chip/array analysis, epidemiology of viral infections, patient diagnostic image analysis (AMI).



The Grid Creates Entirely New Possibilities for Research and Diagnostics!

Erasmus Computing Grid

Erasmus MC

HOGESCHOOL ROTTERDAM

ECG - Business Concepts

The ECG has four major business concepts: the research, education, diagnostics and industry arenas. For all of these, separate business concepts, work-flows and user managements are necessary and need to be integrated to allow maximum efficiency.

The ECG is one of the few grids on this level of complexity !

Research:

- genomic and proteomic analysis
- epidemiology
- image analysis, e.g. Applied Molecular Imaging (AMI)

Education:

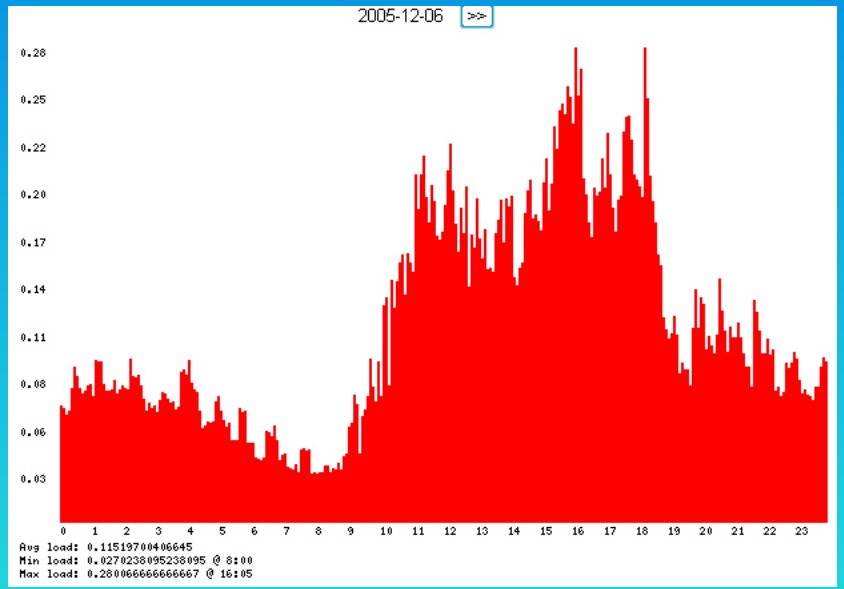
- training of the coming grid generation of IT specialists
- developing new concepts for grid computing

Diagnostics:

- clinical image and data analysis
- operation planning and operation support

Industry:

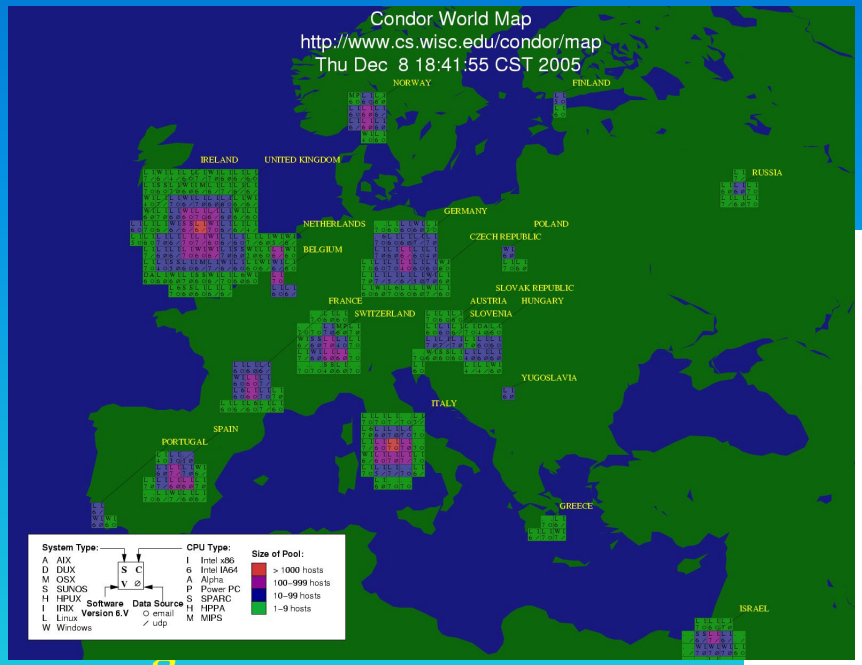
- brokerage of computing resources

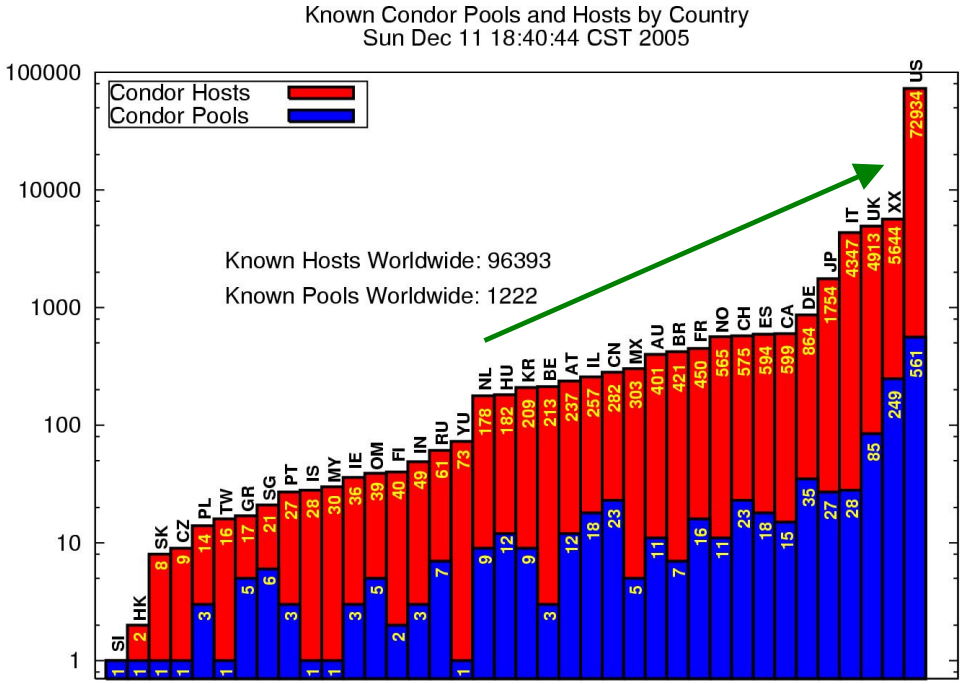


ECG - Business Concepts

The ECG has four major business concepts: the research, education, diagnostics and industry arenas. For all of these, separate business concepts, work-flows and user managements are necessary and need to be integrated to allow maximum efficiency.

The ECG is one of the few grids on this level of complexity !





- clinical image and data analysis
- operation planning and operation support

Industry:

- brokerage of computing resources

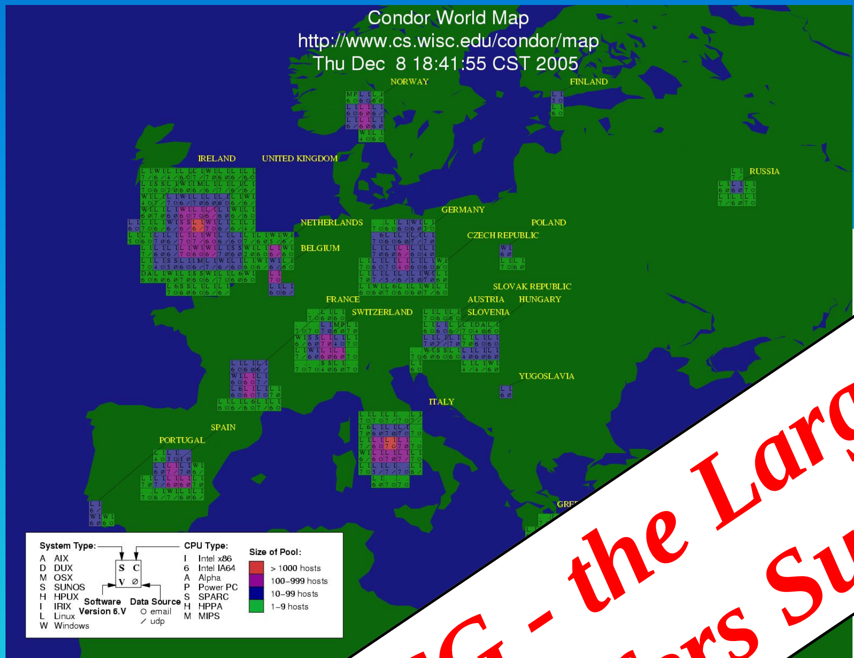
ECG - Business Concepts

The ECG has four major business concepts: the research, education, diagnostics and industry arenas. For all of these, separate business concepts, work-flows and user managements are necessary and need to be integrated to allow maximum efficiency.

The ECG is one of the few grids on this level of complexity !

Erasmus Computing Grid

HOGESCHOOL ROTTERDAM

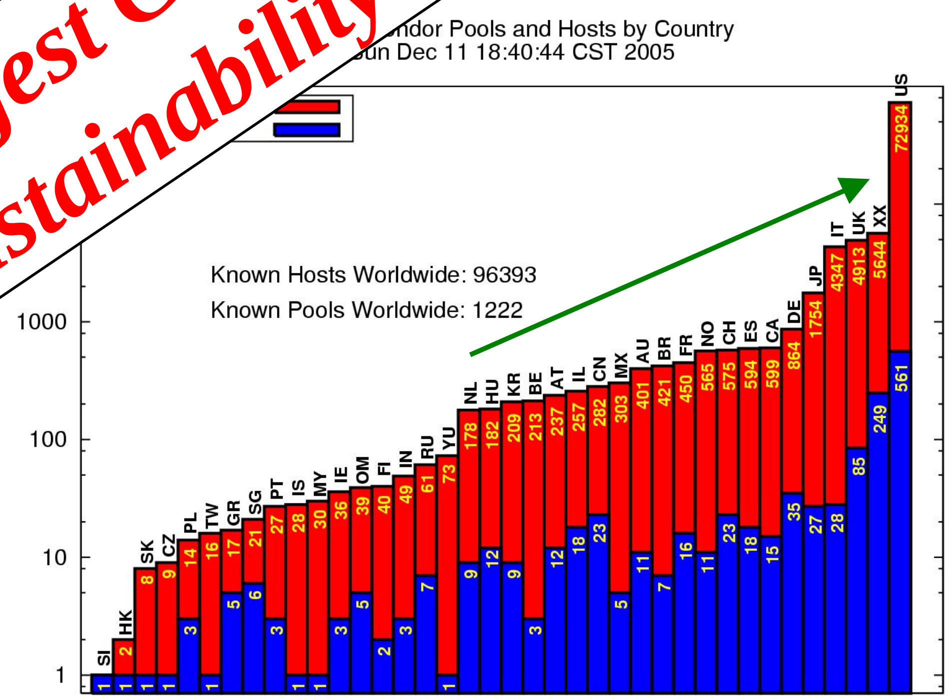


The ECG - the Largest Condor Grid World-Wide - Offers Sustainability & Security !

- clinical i...
- ... support

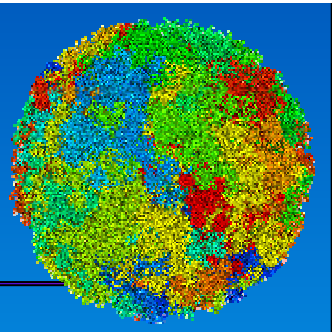
In...

- bi... computing resources



The 3D Genome Viewer

To develop an entire novel system-biology oriented genome browser, i.e. integrating the holistic complexity of genome organization in a single easy comprehensible platform has required completely new approaches to represent the genome architecture realistically in combination with the various types experimental data or instant analysis capabilities and annotation.



Paradigm:

- **simplicity, flexibility, customisability, expandability and open-source sustainability**

  concerning

- **user accessibility and needs, the in-/output of data sources, system modularity, and integration of future developments**

Content:

- **all spatial scales from the base-pair to nuclear morphology including all other system composites**
- **every type of even large multi-dimensional datasets**

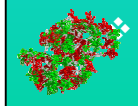
User Handling:

- **intuitive navigation with dynamical semantic resolution and relational arrangement**

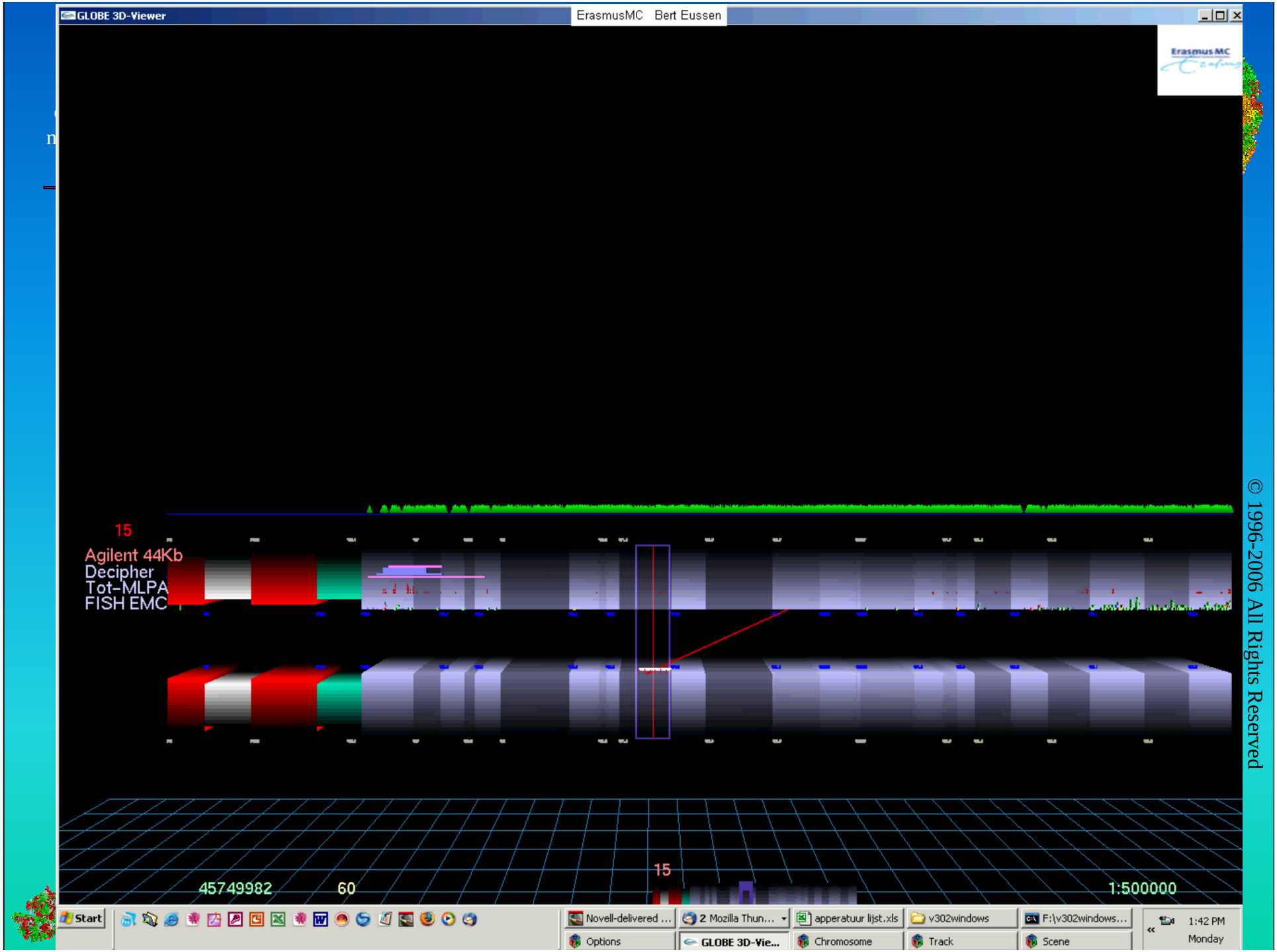
Data Handling:

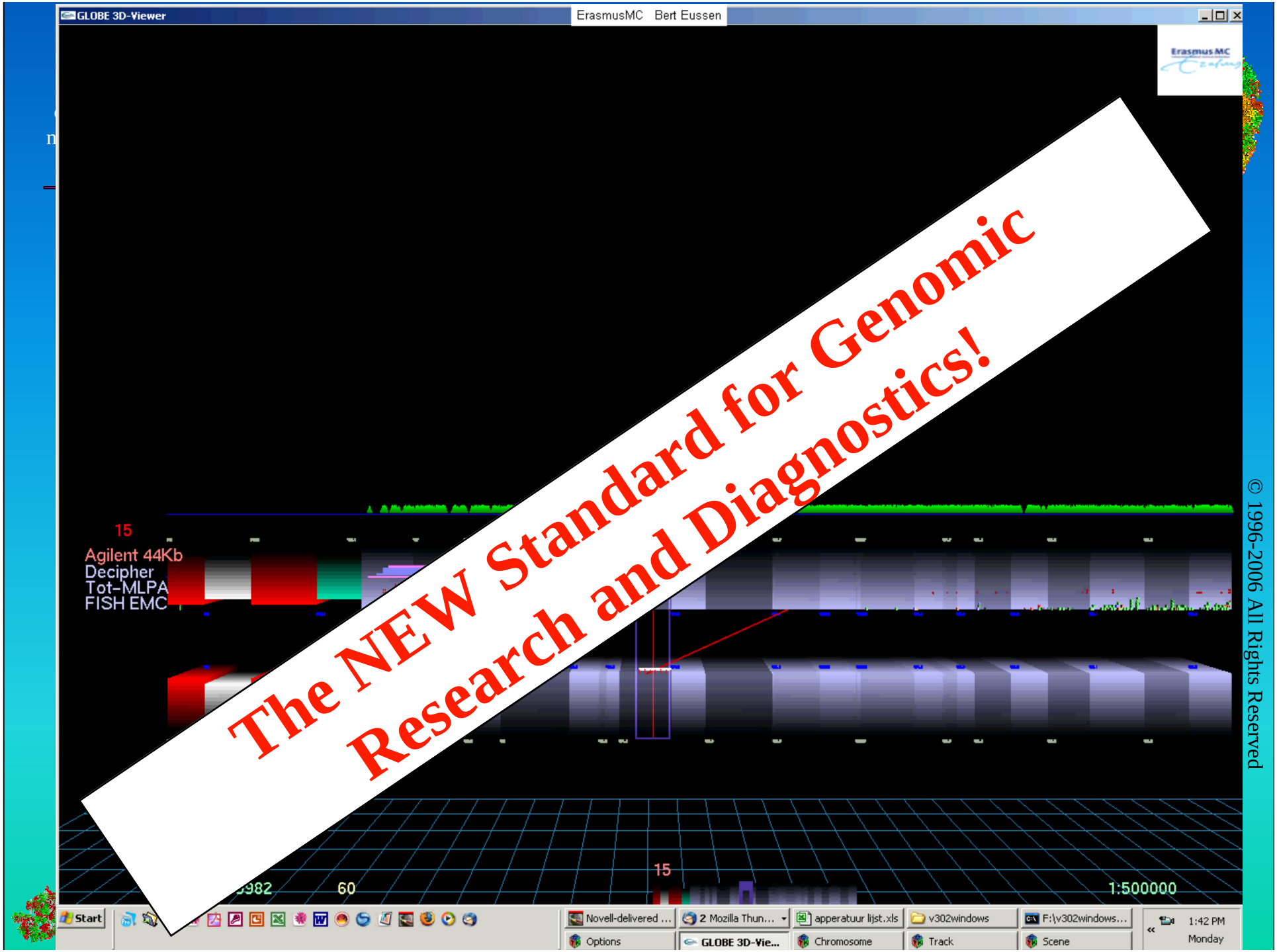
- **real-time interaction with data manipulation or annotation**
- **simple relational or complex correlative multi-dimensional analysis and planning capabilities**

Control Options:

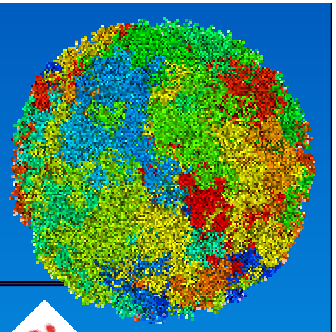


- **front-end access and control of virtual system biological genome simulations**





The 3D Genome Viewer



Due to its broad impact, the GLOBE 3D Genome Viewer addresses different markets and revenue streams in the areas of research, education, diagnostics and industry with different detailed business models and contracting patterns which will guaranty a constant high-profile valorization.

Basic Corporate License:

- 50% science
- 50% industry

White Label Products:

- special customer products

Content Sale

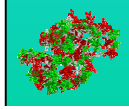
- specially defined data sets
- specially connected data sets

Advertising:

- sponsored free version
- content and user related

eCommerce:

- community portal and referral fees



Companies/Organizations:

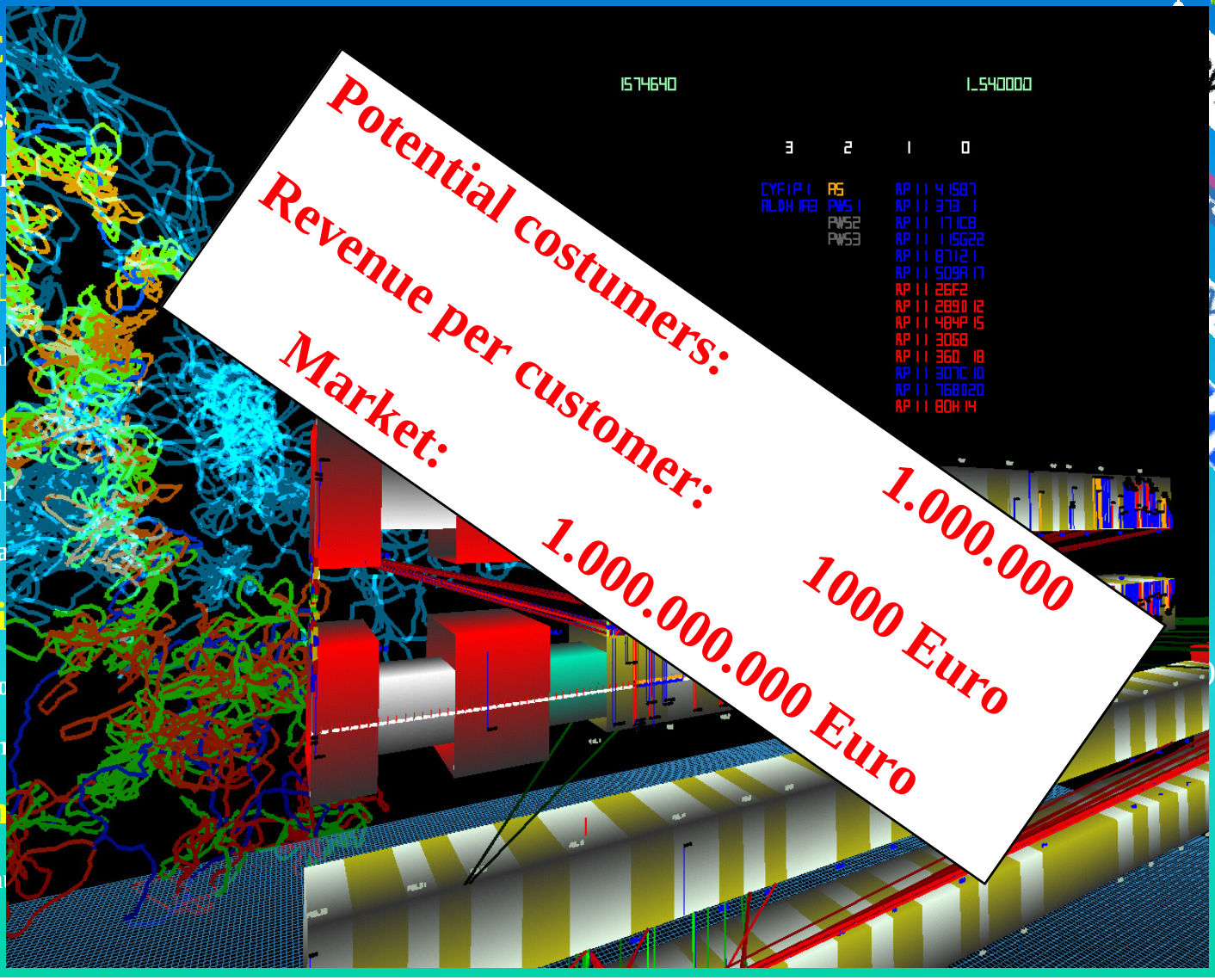
- Invitrogen (Karlsbad, CA, USA)
- Affimetrix (Santa Clara, CA, USA)
- Soft Genetics (State College, PA, USA)
- Combimatrix CMDX (San Antonio, CA, USA)
- Philips (The Netherlands)
- MediGRID (Goettingen, Germany)

Collaborators:

- Prof. Dr. Evan E. Eichler (Seatle, WA, USA)
- Prof. Dr. Stephen W. Scherer (Toronto, Canada)
- Prof. Dr. Michael Hausmann (Heidelberg, Germany)
- Dr. Malte Wachsmuth (Seoul, Korea)
- Dr. Gregory Cox (Karlsbad, CA, USA)
- Dr. Eberhard Schmitt (Jena, Germany)
- Dr. Sabine Baars (Leiden, The Netherlands)

The 3D Genome Viewer

Due to its broad impact, the GLOBE 3D Genome Viewer addresses different markets and revenue streams in the areas of research, education, diagnostics and industry with different detailed business models and contracting patterns which will guaranty a constant high-profile valorization.



Potential costumers: 1.000.000

Revenue per customer: 1000 Euro

Market: 1.000.000.000 Euro

Basic C

- 50% s
- 50% i

White L

- special

Content

- special
- specia

Adverti

- sponso
- conten

eComm

- commu

The 3D Genome Viewer

Due to its broad impact, the GLOBE 3D Genome Viewer addresses different markets and revenue streams in the areas of research, education, diagnostics and industry with different detailed business models and contracting patterns which will guaranty a constant high-profile valorization

Basic C

- 50% s
- 50% i

White I

- special

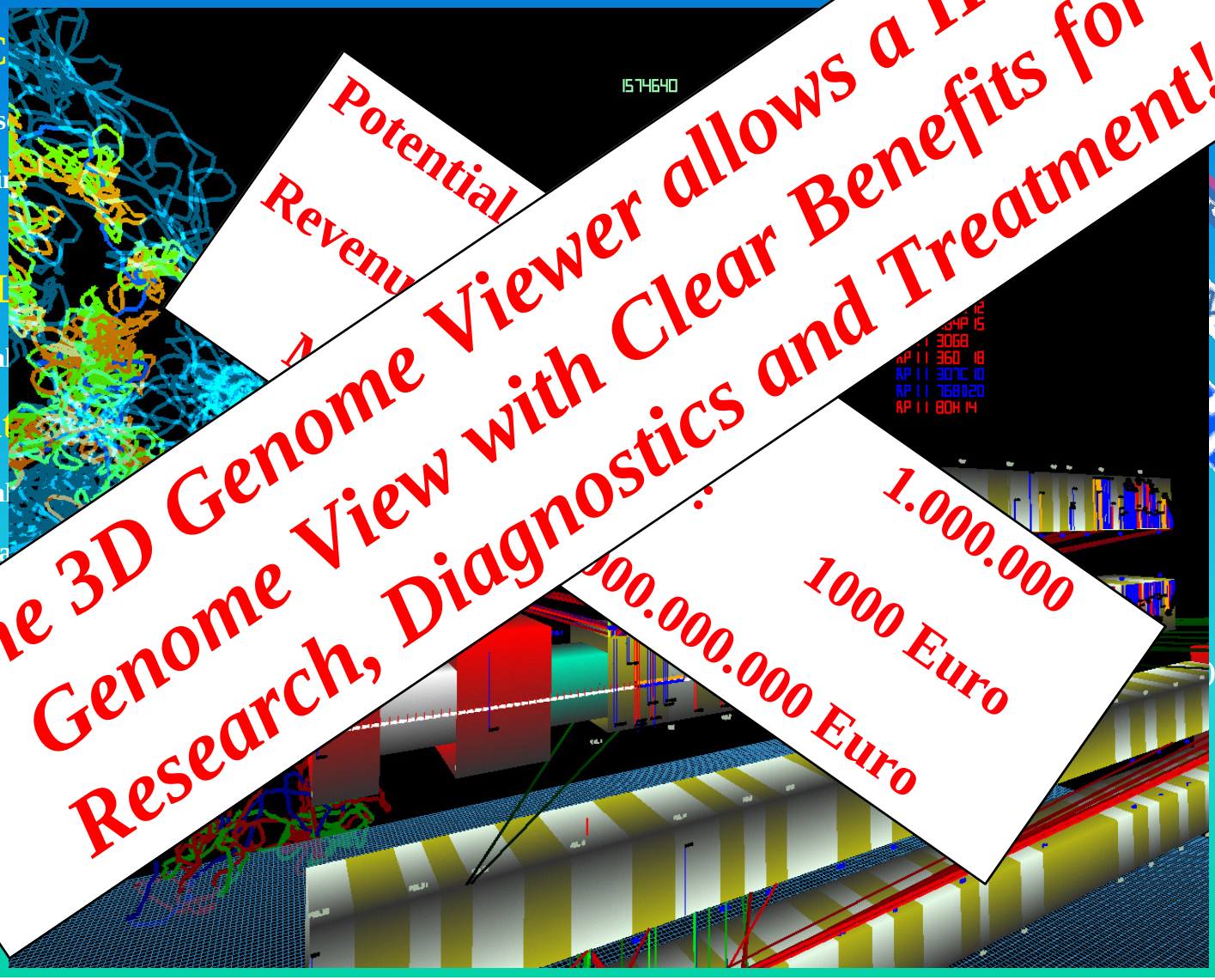
Content

- special
- specia

Ad

eC

- co



Potential

Revenu

N

1.000.000

1000 Euro

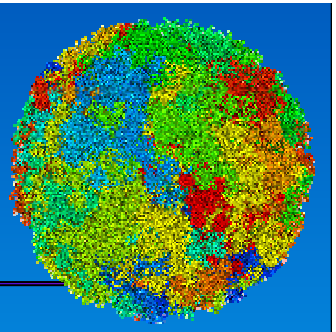
000.000.000 Euro

The 3D Genome Viewer allows a Holistic Genome View with Clear Benefits for Research, Diagnostics and Treatment!

The 3D Genome Viewer

The NEW Standard for Genomic Research and Diagnostics!

Ready to be Valorized!!!



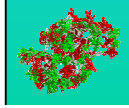
2007 2008 2009

Basic Corporate Licenses

White Label Products & Advertising

e Commerce

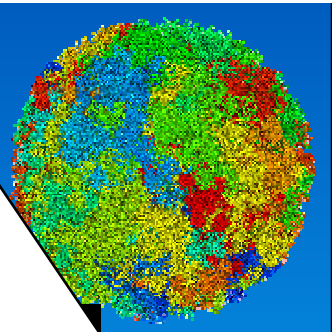
Content Sale



The 3D Genome Viewer

The NEW Standard for Genomic Research and Diagnostics!

Ready to be Valorized!!!



2007

Basic Co

The 3D Genome Viewer allows a Holistic Genome View with Clear Benefits for Research, Diagnostics and Treatment!

Opportunities

We have started to establish successfully a unique environment for genomic research and diagnostics leading to a holistic understanding of genomes necessary for the improvement of mankind!

The Erasmus Computing Grid:

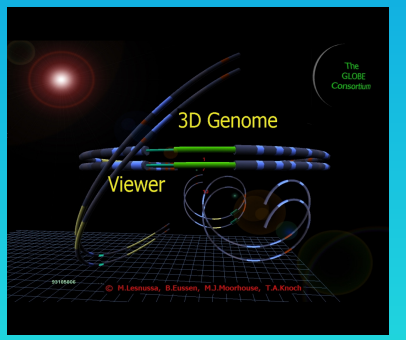
The Erasmus Computing Grid greatly advances the computing capabilities of the Erasmus Mc, e.g. for:

- **genomic and proteomic analysis**
- **epidemiology**
- **clinical image analysis, e.g. Applied Molecular Imaging (AMI)**

The 3D Genome Viewer:

The 3D Genome Viewer greatly advances the integration of multi-dimensional data, e.g.:

- **sequential genomic data**
- **structural genomic data**
- **clinical diagnosis data**



Both the Erasmus Computing Grid and the 3D Genome Viewer have great potential for various forms of commercialization!

Opportunities

We have started to establish successfully a unique environment for genomic research and diagnostics leading to a holistic understanding of genomes necessary for the improvement of mankind!

The Erasmus Computing Grid:

The Erasmus Computing Grid greatly advances the comput

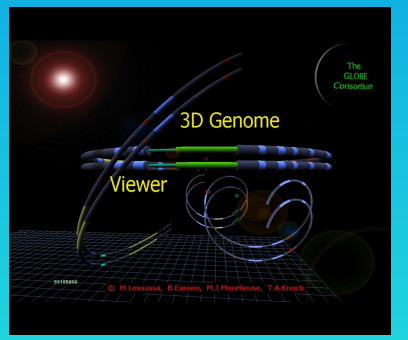
- **genomic and proteomic analysis**
- **epidemiology**
- **clinical image analysis, e.g. Appli**

The 3D Genom

The 3D Geno **ulti-dimensional data, e.g.:**

-

Both the Erasmus Computing Grid & the 3D Genome Viewer have great potential for commercialization!



asmus Computing Grid and the 3D Genome Viewer reat potential for various forms of commercialization!

Information Life-Cycle Management at the Erasmus Medical Center

-

Collaboratively Managing Digital Data for Care, Research, Education

and the

International Development of the GLOBE 3D Genome Viewer

Knoch, T. A., Walgemoed, P. & Eussen, B.

2nd Annual World Healthcare Innovation and Technology Congress, Washington DC, USA, 1st - 3rd November, 2006.

Abstract

Information Lifecycle Management at the Erasmus University Medical Centre. Collaboratively managing digital data for care, research and education using the international development of the GLOBE 3D Genome Viewer and Erasmus Computing Grid as catalyzing initiatives.

The Rotterdam-based Erasmus MC is with 1,200 patient beds and 10,000 employees the largest biomedical research and healthcare institution in the Netherlands and one of the biggest in Europe. Erasmus MC witnessed a steep increase in data volumes and subsequent computing and archiving needs for its care, research and education processes. The continuous advances of healthcare imaging and other data in combination with the unusually long archiving lifecycles propose a complex availability challenge. Addressing this challenge requires a collaborative interdisciplinary effort of Clinical, Research, Education and IT departments across Erasmus MC. Fulfilling the healthcare vision of the 21st century for a higher quality of life in our societies necessitates the crossing of the institutional border by alliances between international hospital partners and industry vendors.

The **GLOBE 3D Genome Viewer** and the **Erasmus Computing Grid** are two new major catalyzing initiatives at Erasmus MC towards achieving this 21st century vision. The *GLOBE 3D Genome Viewer* is the first system-biology oriented genome viewer necessary to access, present, annotate, and simulate the holistic genome complexity in a unique gateway towards a real understanding, educative presentation and curative manipulation planning. The *Erasmus Computing Grid* is with its projected 13,000 CPUs and a capacity of ~20 Tera flops one of the biggest desktop computing grids in the world supplying the computing power needed for research, education and healthcare.

Erasmus MC likes to call for collaboration on the *GLOBE 3D Genome Viewer* for visualizing the tremendously complex genome landscapes and biological networks in a single easy comprehensible platform necessary to achieve the challenge posed by the upcoming of the complete sequencing of individual genomes and thus personalized medicine.

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.