FP7 Support Action - European Exascale Software Initiative

DG Information Society and the unit e-Infrastructures

EESI Final Conference

EESI WG3.4: Life Sciences and Health

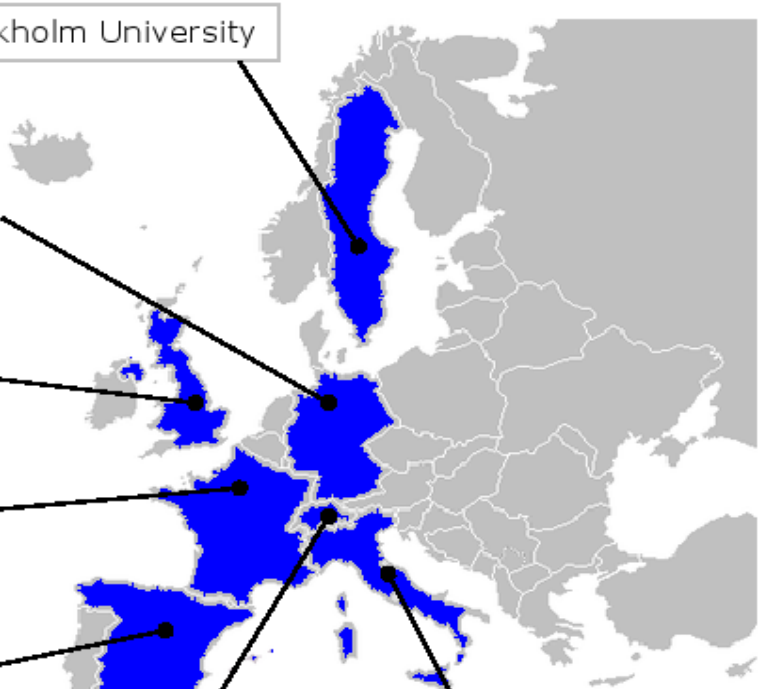
Modesto Orozco, Chair

Janet Thornton, Vice Chair

Ramon Goni, speaker

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

The conclusions of the Life Science panel

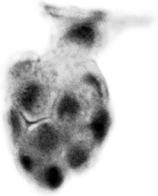


Erik Lindahl	Stockholm University
Helmut Grubmuller	Max Planck Institute
Paolo Carloni	German Research School
Wofgang Wenzel	Karlsruhe Institute of Technology
Reinhard Schneider	EMBL
Janet Thorton/Andrew Lyall	EMBL-EBI
Charles Laughton	University of Nottingham
Richard Lavery	University of Nottingham
Nicolas Baurin	Sanofi-Aventis
Modesto Orozco	BSC
Alfonso Valencia	CNIO
Henry Markram/Felix Schuermann	EPFL
Manuel Peitsch	Swiss Institute of Bioinformatics
Julian.Tirado-Rives (US)	University of Yale
Adrian E. Roitberg (US)	University of Florida
Anna Tramontano	University of Roma

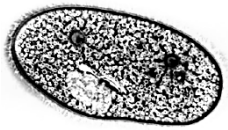
Life Sciences and Health

Population

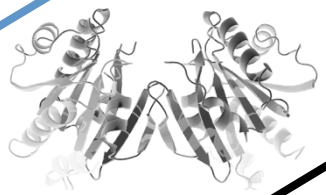
Organ



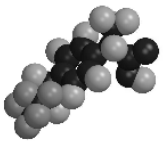
Tissue



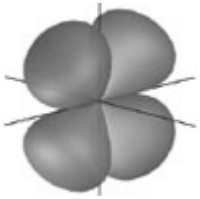
Cell



Macromolecule



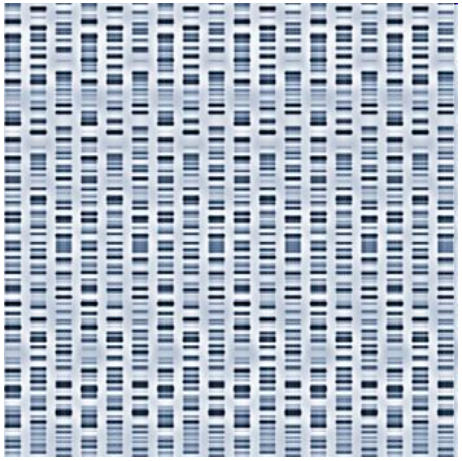
Small Molecule



Atom

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

2011 a unique opportunity

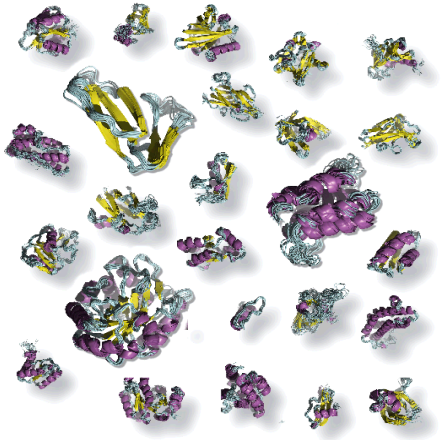


Genomes

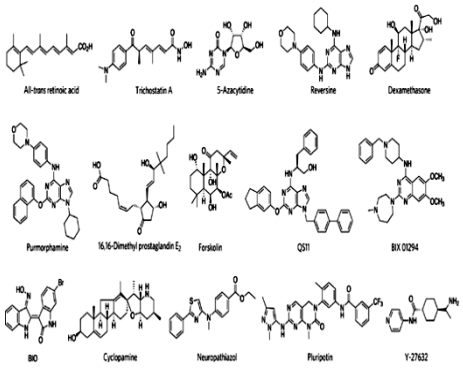
Integration!

Data

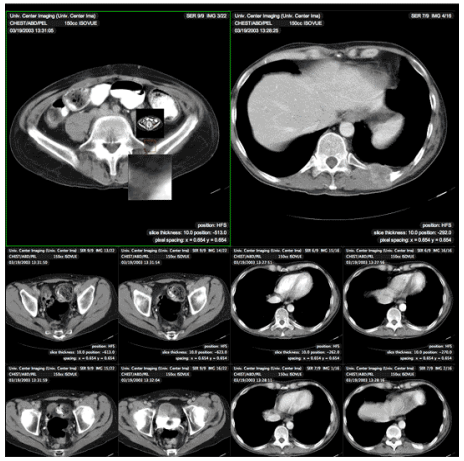
Flops



Structures



Molecules



Images

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

human genome project - Buscar con Google - Windows Internet Explorer

http://www.google.es/search?q=human+genome+project&um=1&hl=es&biw=984&bih=548&sou

Archivo Edición Ver Favoritos Herramientas Ayuda EPSON Web-To-Page

Favoritos IRB Barcelona Mail - Inbox -... Galería de Web Slice Sitios sugeridos

human genome project - Buscar con Google

La Web Imágenes Vídeos Mapas Noticias Libros Correo Más

Google human genome project

Aproximadamente 2.080.000 resultados (0,26 segundos)

Todo
Imágenes
Vídeos
Noticias
Más

Barcelona
Cambiar ubicación

La Web

$1,688 whole human exome
$1,688/1st sample, $1,999 for bulk HiSeq2000(2x100)&NimbleGen v2 capt
www.otogenetics.com

Artículos académicos para human genome project
... expressed sequence tags and human genome project - Ada
1836
New goals for the US human genome project: 1998- ... - Collins
Implications of the Human Genome Project for medical ... - Collin

Human Genome Project Information - [Traducir
3 Feb 2011 ... The main homepage for Human Genome Pro
project is; its progress, history, and goals; what issues are as

ScienceDaily®

Your source for the latest research news

Are you ready to learn how to make a difference?

News Articles Videos Images Books

Health & Medicine Mind & Brain Plants & Animals Earth & Climate Space & Time Matter &

Science News

Blog Cite

Human Genome's Breaking Points: Genetic Sequence of Large-Scale Differences Between Human Genomes

ScienceDaily (Feb. 2, 2011) — A detailed analysis of data from 185 human genomes sequenced in the course of the 1000 Genomes Project, by scientists at the European Molecular Biology Laboratory (EMBL)

enlarge

Chinese bioscience: The sequence factory : Nature News - Windows Internet Explorer

http://www.nature.com/news/2010/100303/full/464022a.html

Archivo Edición Ver Favoritos Herramientas Ayuda EPSON Web-To-Page Imprimir Vista preliminar

Favoritos IRB Barcelona Mail - Inbox -... Galería de Web Slice Sitios sugeridos

Chinese bioscience: The s... IRB Barcelona Mail - Inbox (2... Página Seguridad Herramientas

nature news

Login

nature news home | news archive | specials | opinion | features | news blog | nature journal

comments on this story

Published online 3 March 2010 | *Nature* **464**, 22-24 (2010) | doi:10.1038/464022a

News Feature

Chinese bioscience: The sequence factory

Stories by subject
- Biotechnology
- Genetics
- Developing world

Stories by keywords
- China
- Genomics
- Genome
- Sequencing

This article elsewhere
Blogs linking to

most recent | commented

- Greenland ice-melt map gets the cold shoulder
  20 September 2011
- Texas prepares to fight for stem cells
  2011
- ...tes disgraced doctor
  2011
- ...arths spark desire for a
  2011
- Glowing cells guide cancer surgeons
  18 September 2011

Related stories
- Palaeogenetics: Icy resolve
  10 February 2010
- Genome reveals panda's carnivorous side
  11 December 2009
- Silkworm genome gets solid coverage

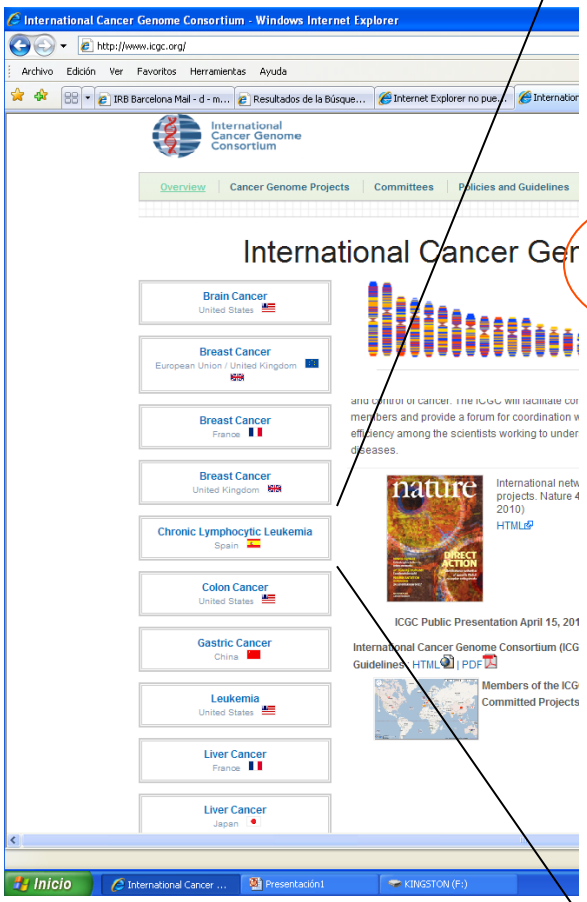
DISCOVER your nature
Buy one-year get two

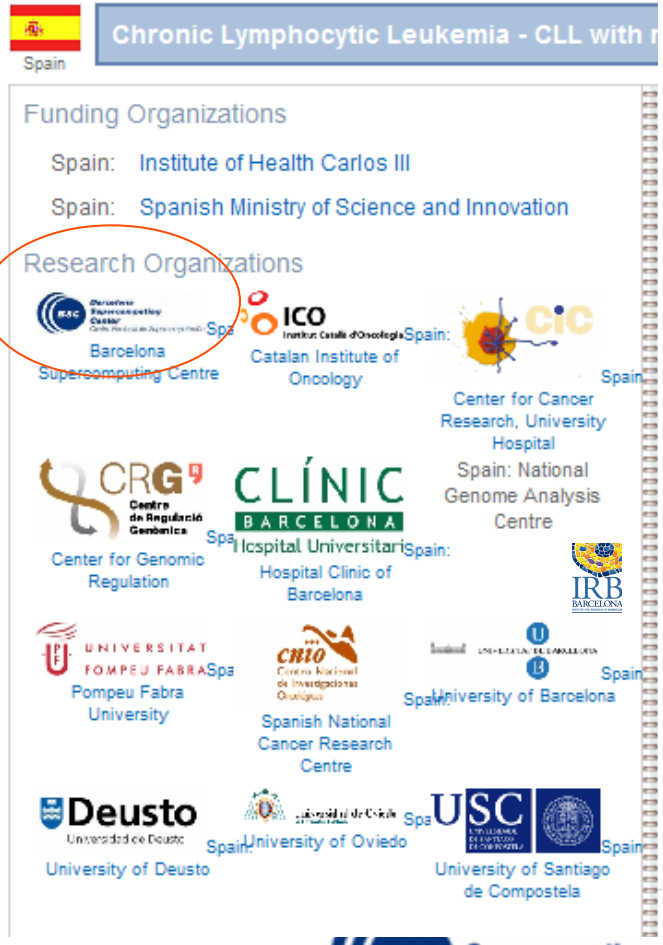
October 2010.
« Although far from comprehe
the tally indicates that at least
human genomes will have bee
completed by the end of this n
and that the total will rise to m
than 30,000 by the end of 201

Sequencing the genome of one million citizenships from China

Internation Cancer Genome Project

Experimental data 1 patient < 1 day
Primary analysis 1 patient → 5 days!
Secondary analysis 1 patient → months!





50 cancers

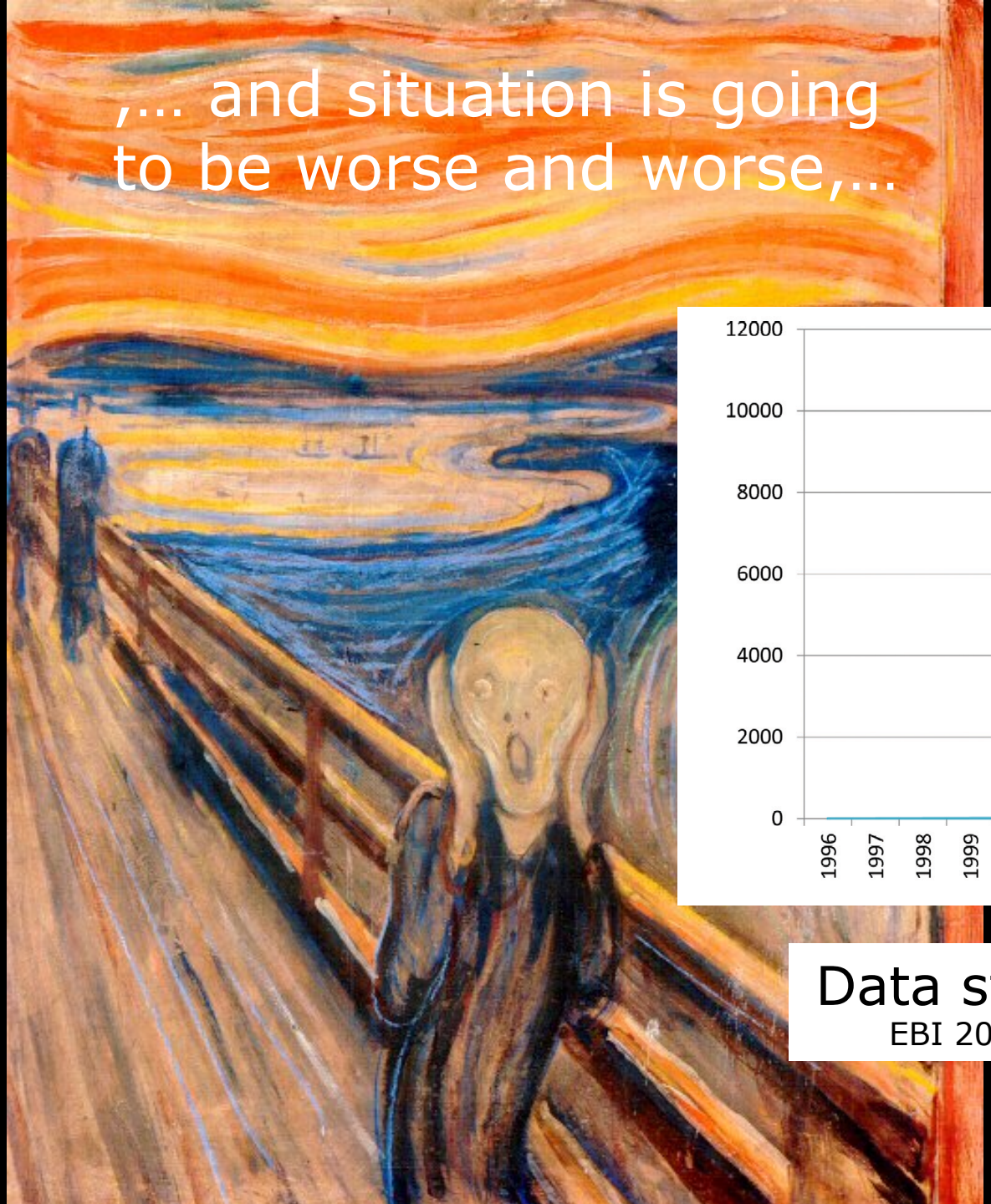
25000 cancer genomes

Puente et al., Nature 2011

Supercomputing Center
Centro Nacional de Supercomputación

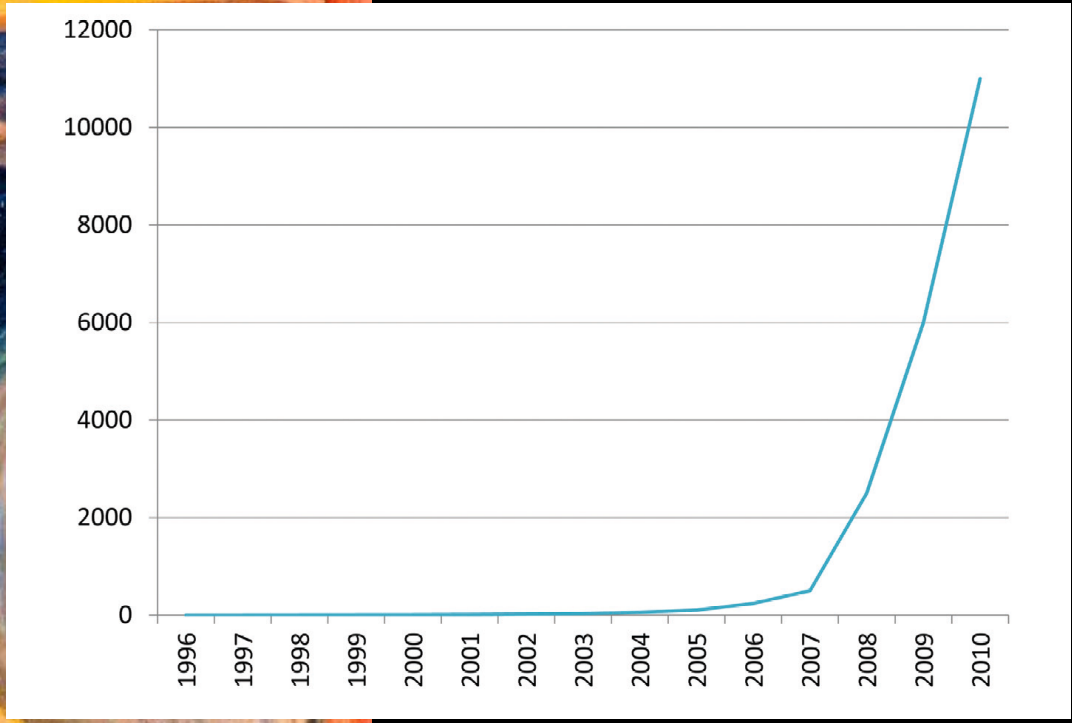
What does it mean,…?

- In BSC experience CLL project:1.5 Pb clean data.
  - We have only 500 patients
  - Analysis means to cross with genomic data for thousands of controls!
  - Data must be confidential, no solutions out of HPC
- CLL implies a sustained use of largest BSC super-computer.
- Computers are becoming the rate-limiting step.



,… and situation is going to be worse and worse,…

ELIXIR

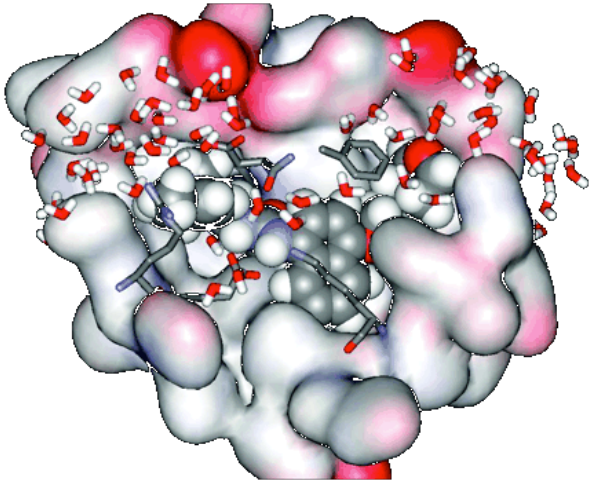


Data storage at EBI (Tb)

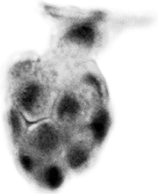
EBI 2010 annual scientific report

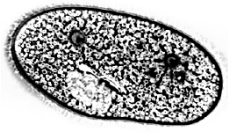
Our dream: simulating life,…

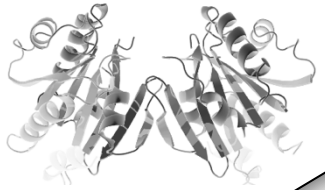
Tremendous impact in reducing the 1.2 Billion $[1] cost of generating a new drug

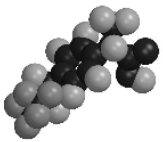


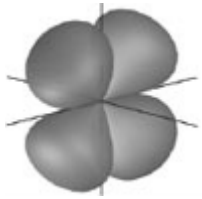
Rational Drug Design











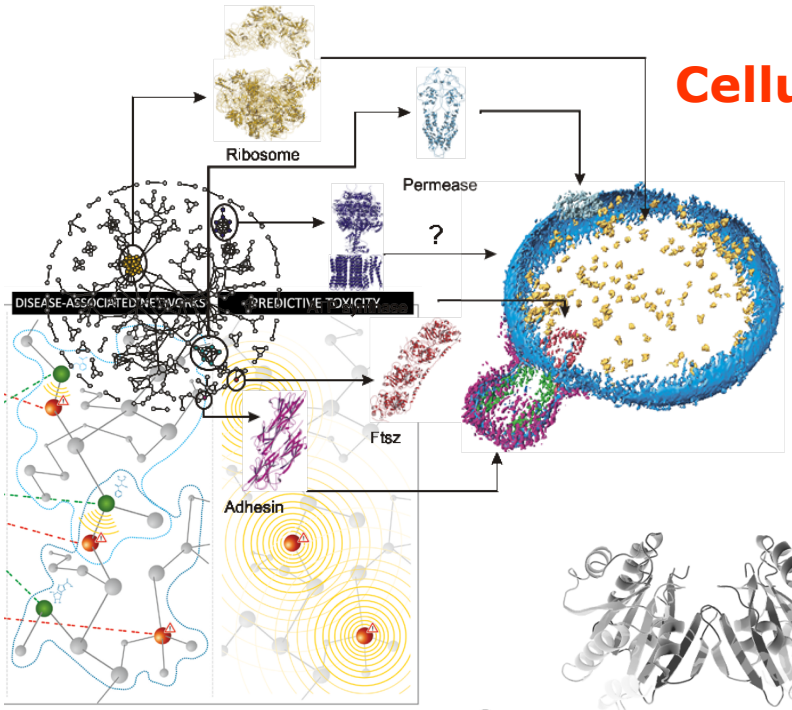
Around 10^5 targets
More than 10^7 drugs
More accurate models!

[1]TUFTS University CSDD Outlook 2008

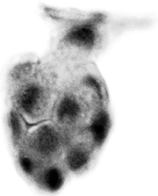
Barcelona Supercomputing Center
Centro Nacional de Supercomputación

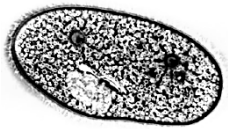
Our dream: simulating life,…

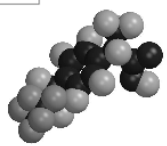
Discovery of new drugs and targets
Reduce side effects and toxicity

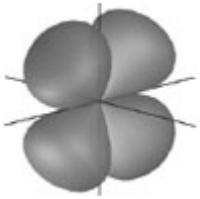


Cellular and subcellular simulation







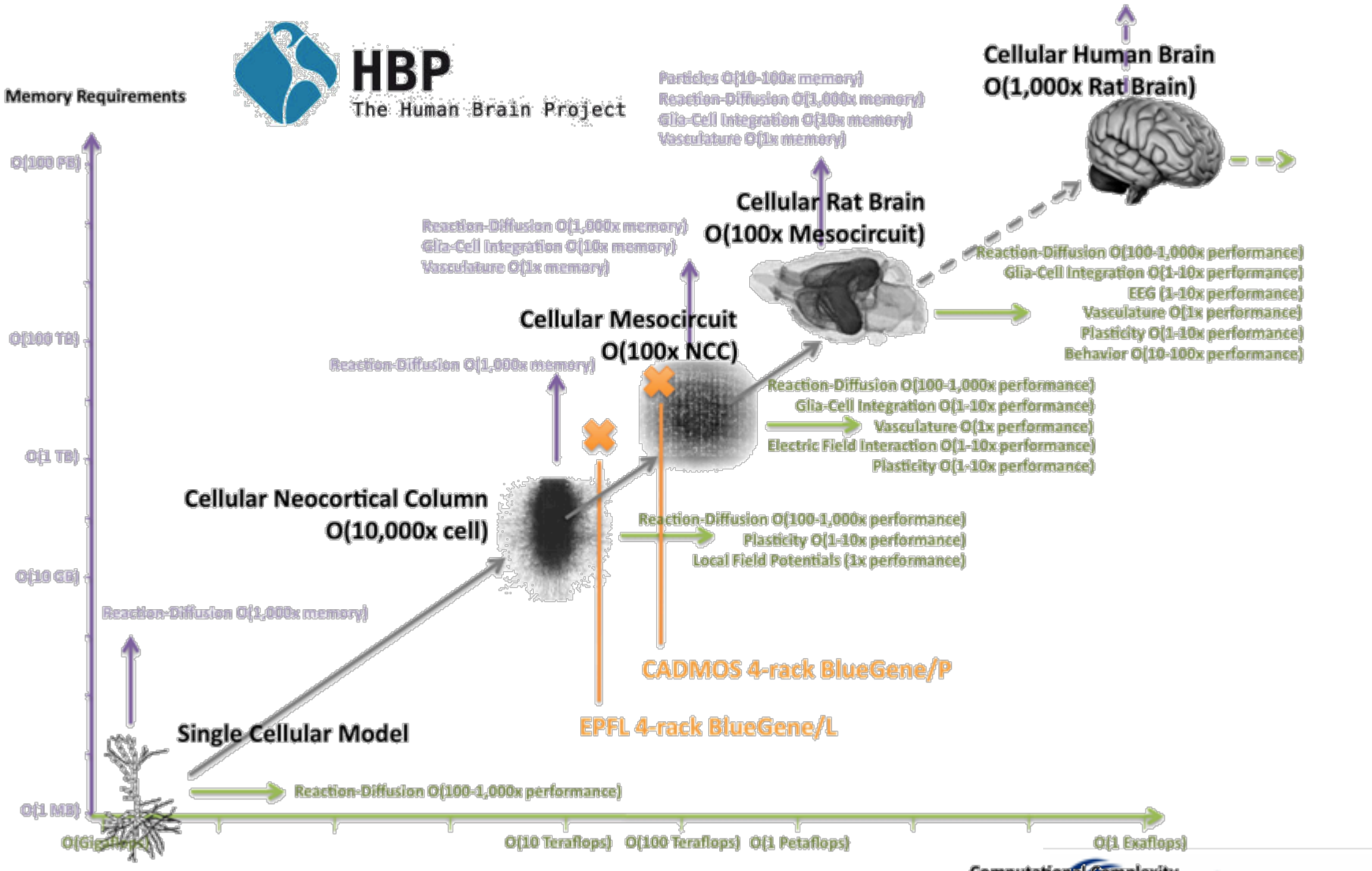


Integration of different data
Need to define models
CPU might be a major issue

Each cell has 10^9 macromolecules

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

FET: Human Brain Project



F Schürmann, H Markram (Blue Brain Project, EPFL)

Genomics and Personalized Medicine

Data is the main problem

ITFoM

FUTURE OF MEDICINE

Treat patients NOT pathologies

Estimated market 450B$ in 2015[1]

[1]2009 PricewaterhouseCoopers

Barcelona Supercomputing Center
Centro Nacional de Supercomputación

Ambition: reinforce Europe position

STRENGTHS

- ELIXIR, Bioinformatics community
- EMBL-EBI
- Research consortia: ENCODE, ICGC, iHEC
- FET Flagships: Personalized Medicine, Human Brain Project.
- PRACE
- Pharmaceutical Industry
- Computational Biology and Bioinformatics Software

WEAKNESSES

- Competitors with privileged access to specialized resources
- No coherent unified data layer available for Bio-data.
- Crucial databases outside Europe.
- Bio-HT techniques are not EU
- Bio-community separated from HPC community

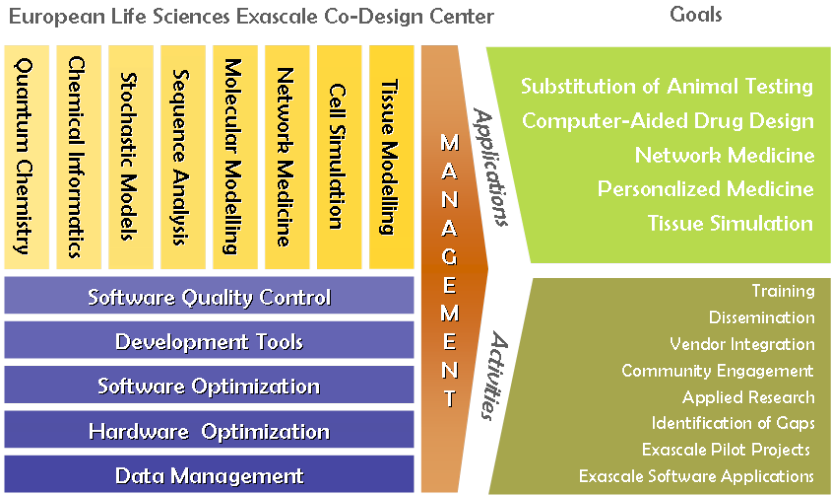
Life-Science needs Exascale

Main conclusions of the panel

- Life science requirements for HPC are enormous and grow exponentially. There is not question that Life Sciences community has ExaScale needs.
- Lack of HPC resources will simply kill several of the most important projects in Life Sciences, including several EU Flagship ones (Human Brain, Personalized Medicine).
- Many other issues rather than FLOPS are important in Life Sciences, often more important.
- **Exascale** should be the keyword, not ExaFlop.
- The experts propose Co-development programs to deploy Exascale to Life Sciences

The Cost

- Estimated around 200 biologists and computer scientist
- Expected investment around 200 MEuros in 2012-2021



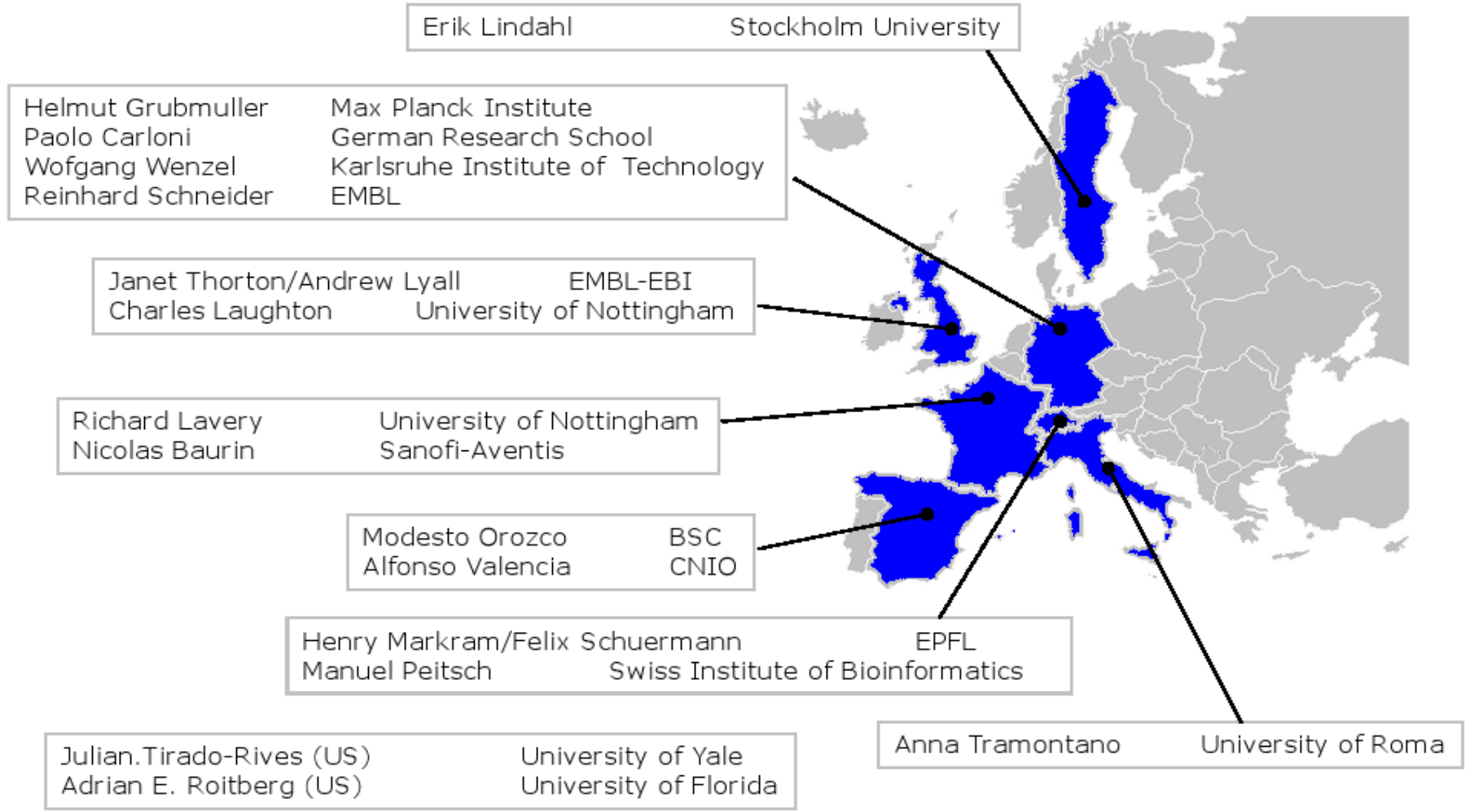
- Public expenditure on healthcare in the EU is around 1,000 B€, the 10% of the GDP.
- The pharmaceutical sector represented a market of roughly 800B€ in 2010
- Europe launches almost the 40% of the pharmaceutical products on the market.
- There are more than 2,100 biotechnology companies in Europe.
- Related areas: cosmetics, food, Environment, Bio-fuels (40% EC Cooperation research budget) will also largely benefit from the initiative.

Impact in health is the crucial issue

One exemple,…

54 million vertebrate animal testing is required to evaluate the safety of major chemicals sold in Europe (see European REACH) Its estimated cost is between 1.3 to 9.5 B€[1].

Animal substitution represents a big reduction of cost of clinical trials by predicting side effect before testing
Development of *in silico* models could simulate entire tissues

[1]http://ec.europa.eu/environment/chemicals/reach/reach_intro.htm

The conclusions of the Life Science panel