

LIFE SCIENCES COMPUTING NEEDS AND THE GRID

DR. ERIK BONGCAM-RUDLOFF

Department of Animal Breeding and
Genetics and

Department of Immunology, Genetics and
Pathology, Uppsala University, Sweden.

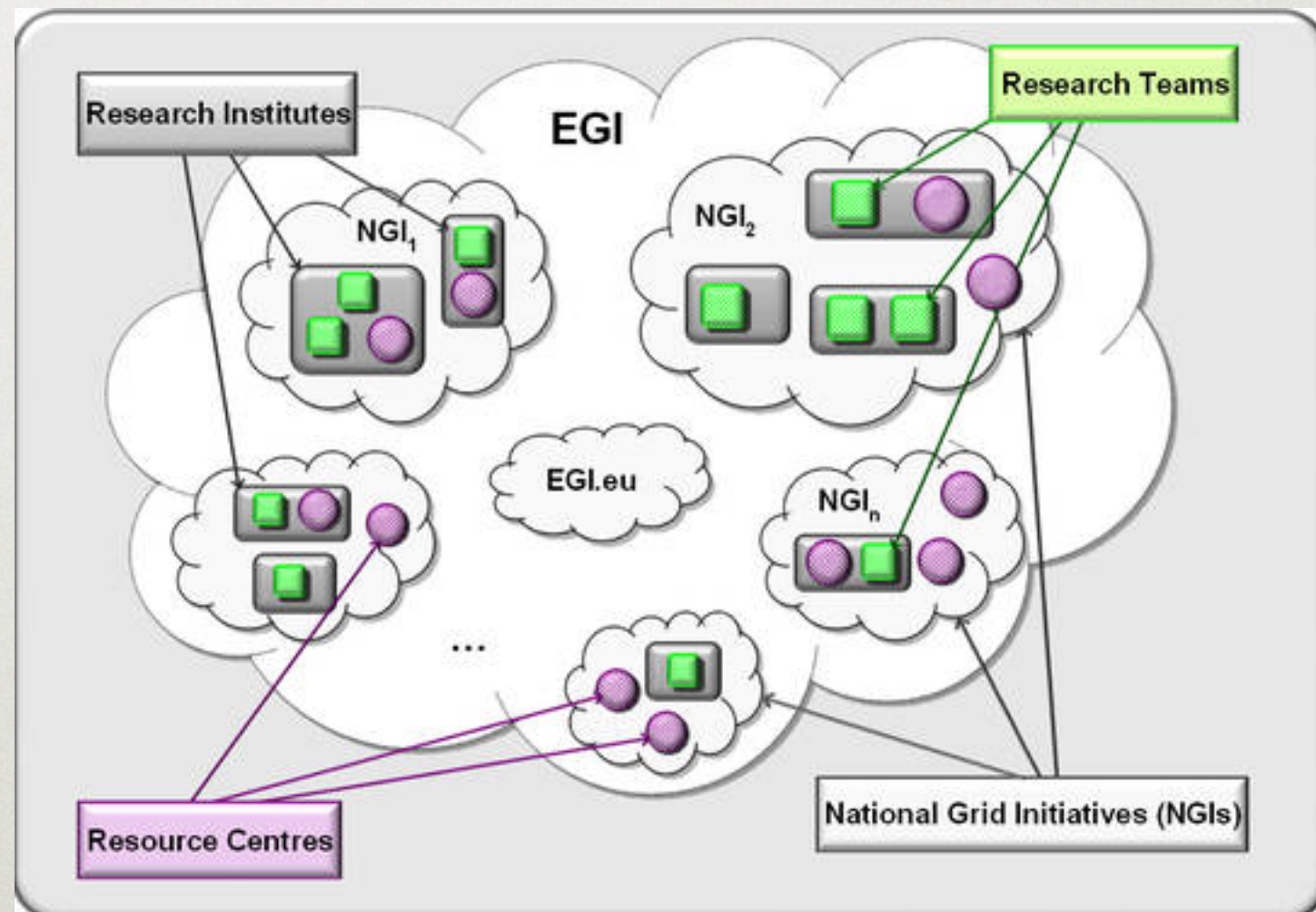


Grids are often differentiated into:

- computational, data
- and collaboration grids.

GRID

- SweGrid, Nordugrid,
EGI



GRID EXAMPLES

- EMBRACE (FP6 NoE): webservices
- BioGrid Australia: connecting patient data
- e-NMR accelerates the prediction of protein structures

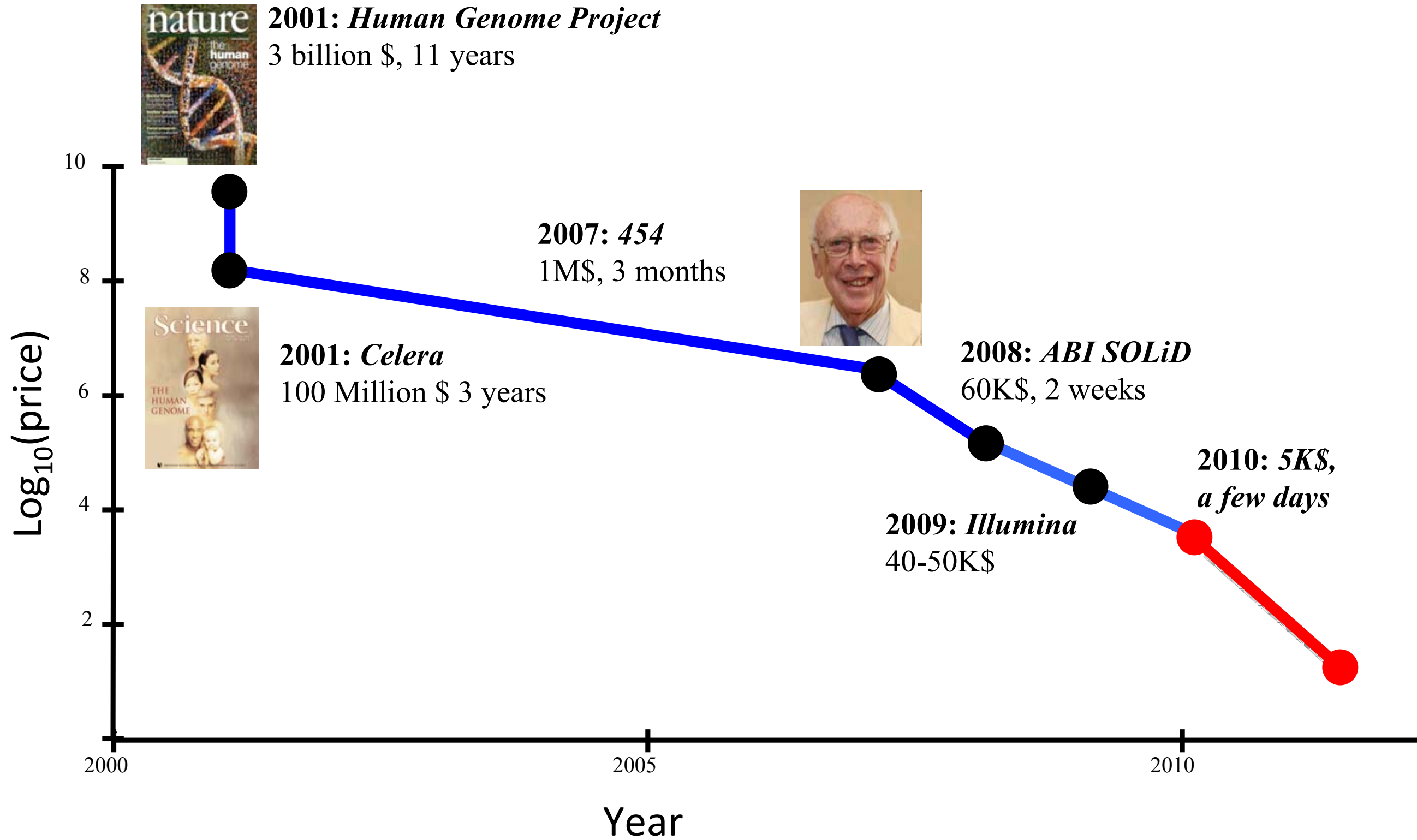
PROTEIN STRUCTURES

- e-NMR accelerates the prediction of protein structures
- Humans have ~25 000 genes
- Each gene has several splice variants
- Therefore several 100 000 different proteins structures could exist

HUMAN GENOME

- 10 years to sequence at high cost
- First draft year 2000
- 25 000 genes= coding
- Coding sequence is only 1% of the genome (as an example ~8 % is rest of viral elements)

Sequencing the Human Genome



Beijing, 31 October–3 November 2010

NEXT GENERATION SEQUENCING

- First
- Second
- Third

The Sequencing Revolution

Next Generation Sequencing

High-Throughput Sequencing
2000



96 sequences per hour

High-Throughput Sequencing
2010



2.6 million sequences per hour

TODAY



ABI SOLiD



454 FLX



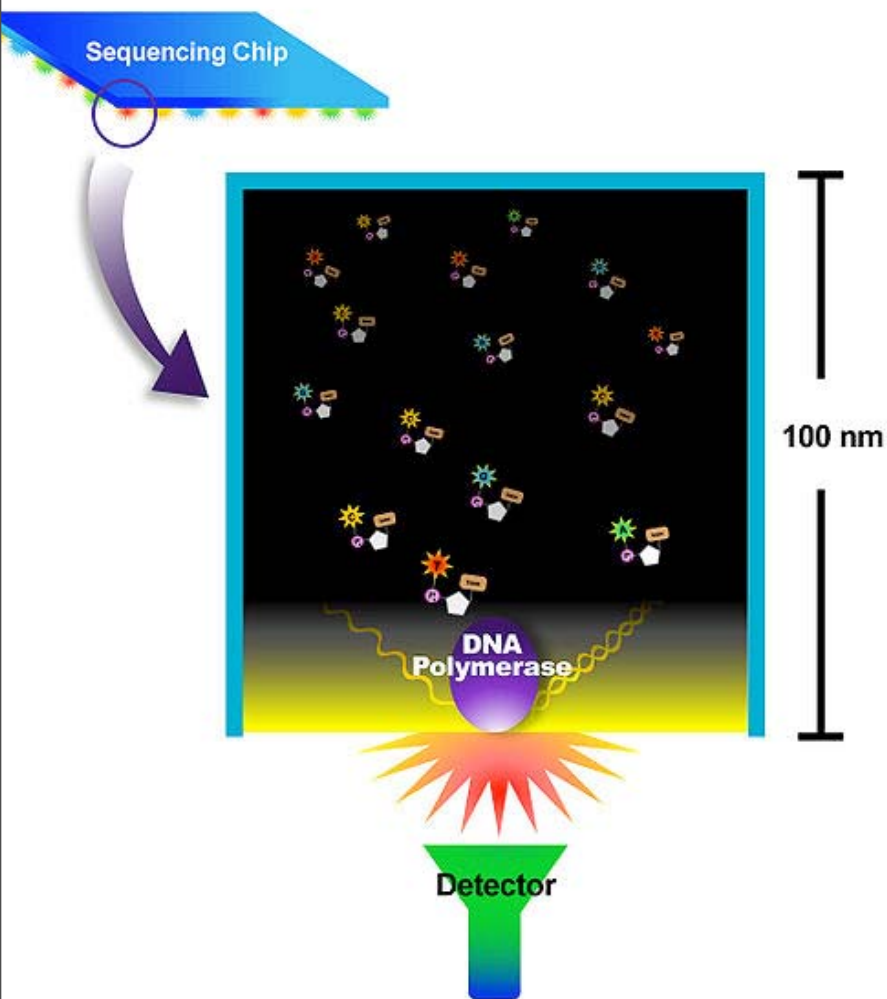
Illumina Genome Analyzer

The current "next-gen" sequencing platforms

Third Generation Sequencing

Single Molecule sequencing

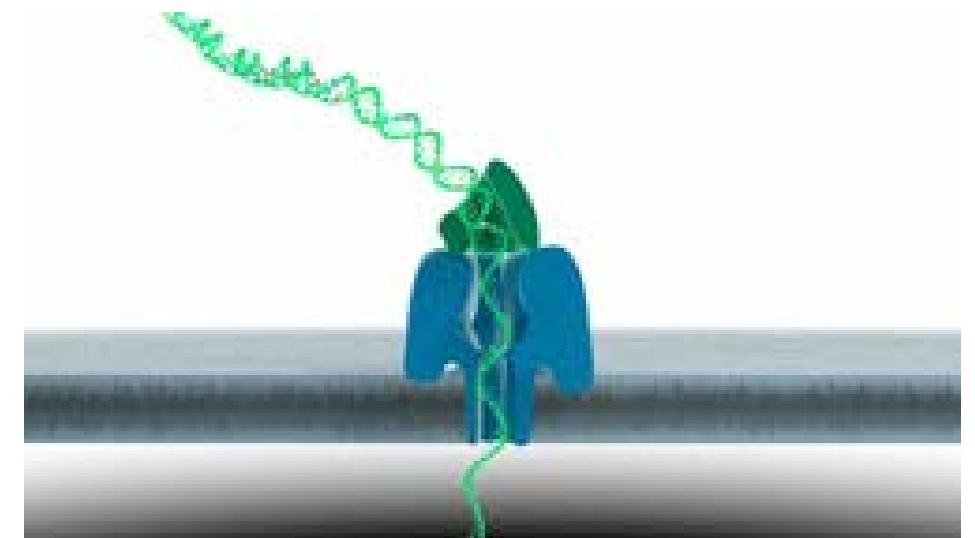
Pacific Biosciences



~3,000 wells per chip
1,500 bp per well
10 bp per second

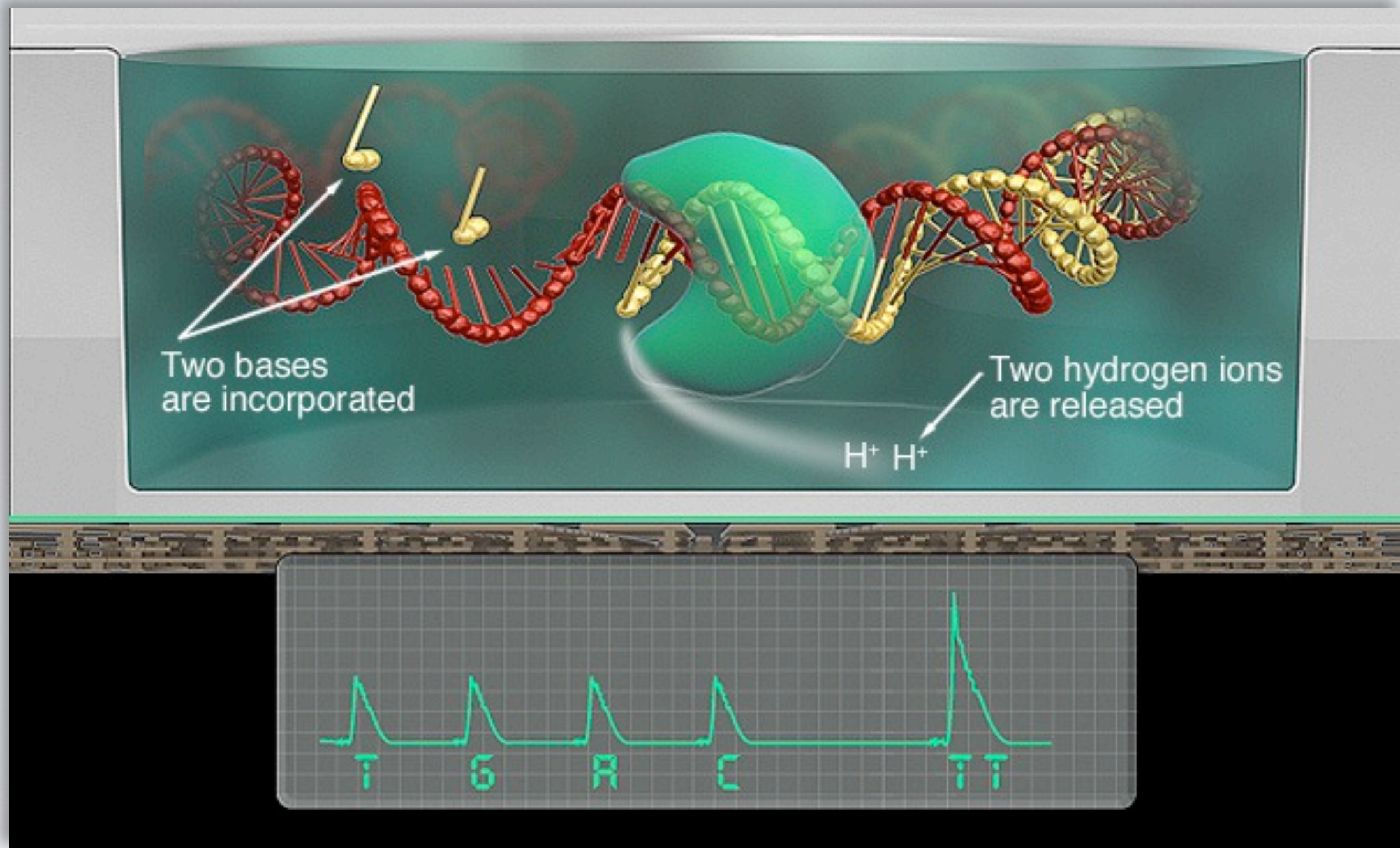
<http://www.pacificbiosciences.com/>

Oxford Nanopore



\$1,000 human genome

IONTORRENT

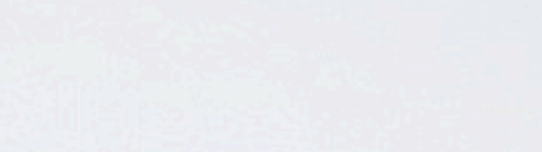


<http://www.iontorrent.com/>

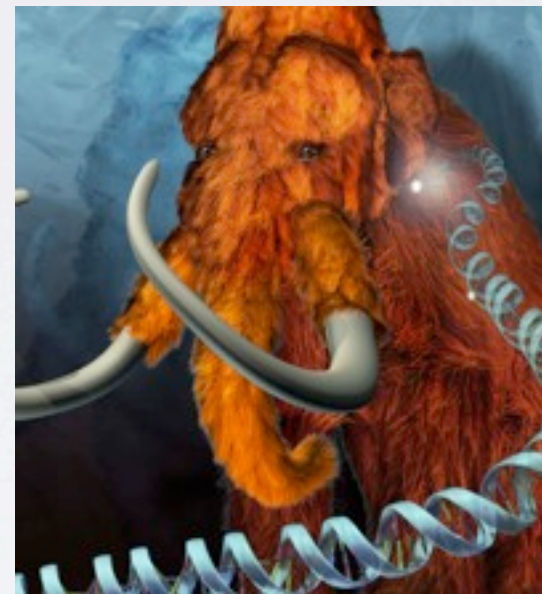
NGS SUCCESS STORIES



Tomato Genome



Mammoth Genome



Chicken domestication



NEXT GENERATION SEQUENCING DATA ANALYSIS NETWORK



<http://www.seqahead.eu/>

<http://www.cost.esf.org>

SCIENCE FICTION?

- No!, things are happening as we speak

NEW E-HOME MEDICINE



23andMe genetics just got personal.

Search 23andMe

Go

Log in

Register Your Kit

Blog

Help

Cart

welcome

ancestry

health

how it works

store

Start filling in the gaps with your DNA



"Because I had given my doctor information from 23andme, he got to a diagnosis much faster. 23andme saved my life." Kirk C.

\$99*

Our new low price for all!
Was \$199

Order Now »

*Requires a 1-year commitment to the [Personal Genome Service®](#) at \$9/mo. **Order for \$399** without commitment.

1 Get Your Kit



2 Provide Saliva



3 Learn About Yourself



4 Get Monthly DNA Discoveries



Gain insight into your traits, from baldness to muscle performance. Discover risk factors for 97 diseases. Know your predicted response to drugs, from blood thinners to coffee. And uncover your ancestral origins. [start tour »](#)

Overview

Discover Health & Ancestry

Keep Your Doctor Informed

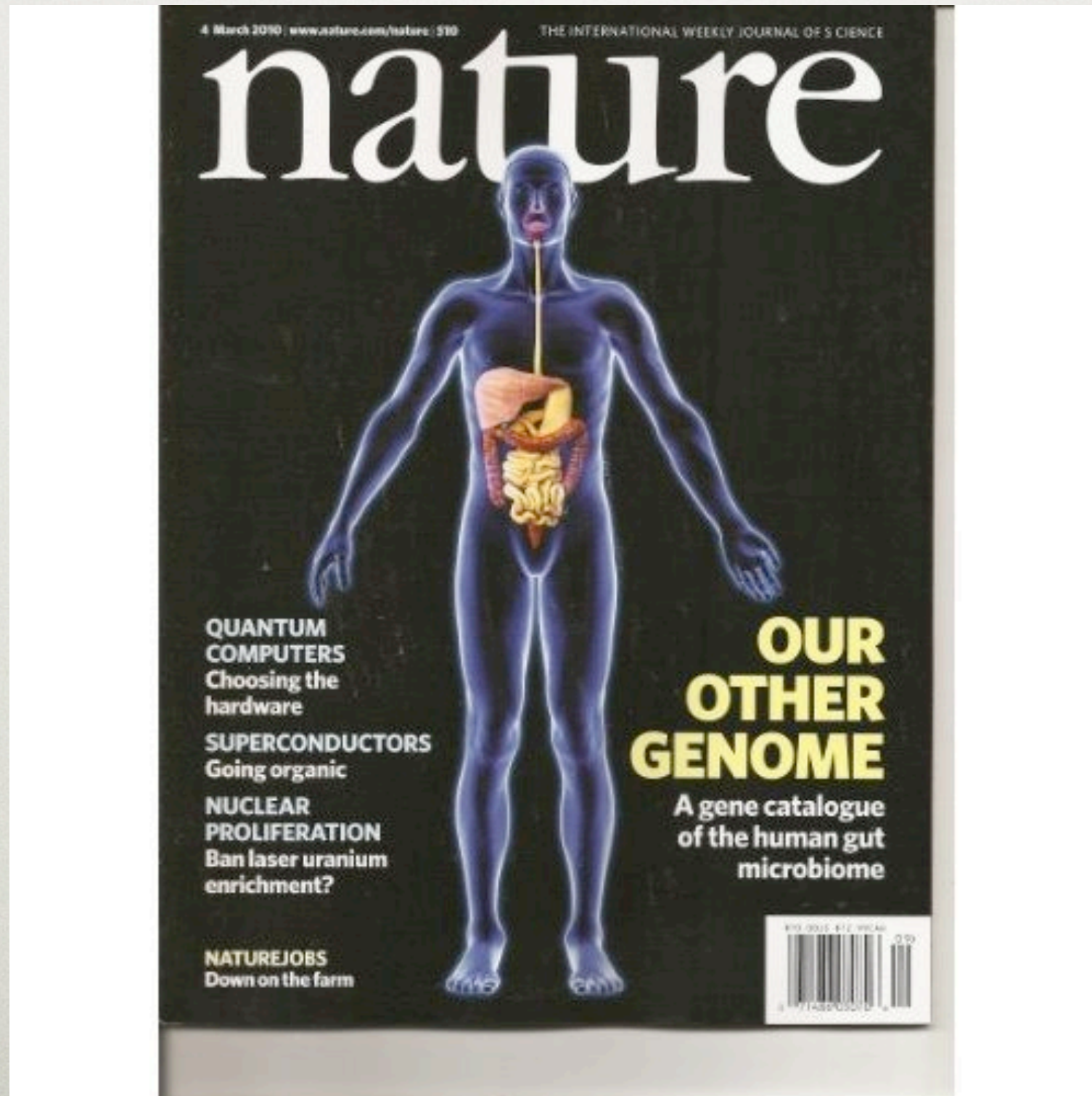
Participate In Research

BIOGRID AUSTRALIA

CELLS

- The average human body, consisting of about 10^{13} (10,000,000,000,000 or about ten trillion) cells
- The human “microbiome” is composed of 500 to 1000 species of bacteria living in the human gut and a roughly similar number on the skin
- Only in the gut the human has ten times more bacterial cells than the human body.

MICROBIOME



METAGENOMICS

- The study of the DNA of uncultured organisms
99% of all microbes can not be cultured
- A GENOME:
Entire genetic information of a single organism
- A METAGENOME:
Entire genetic information of a community of organisms

FUTURE:CONNECT

- Systems like the BioGrid in Australia
- and Metagenomics data
- Do fast analysis and present the results to the patient and doctor in matter of seconds

HOW MANY SPECIES

- According to the Global Taxonomy Initiative and the European Distributed Institute of Taxonomy, the total number of species for some phyla may be much higher as what we know currently:
- 10–30 million insects (of some 0,9 we know today)
- 5–10 million bacteria;
- 1.5 million fungi; (of some 0,4 million we know today)
- ~1 million mites

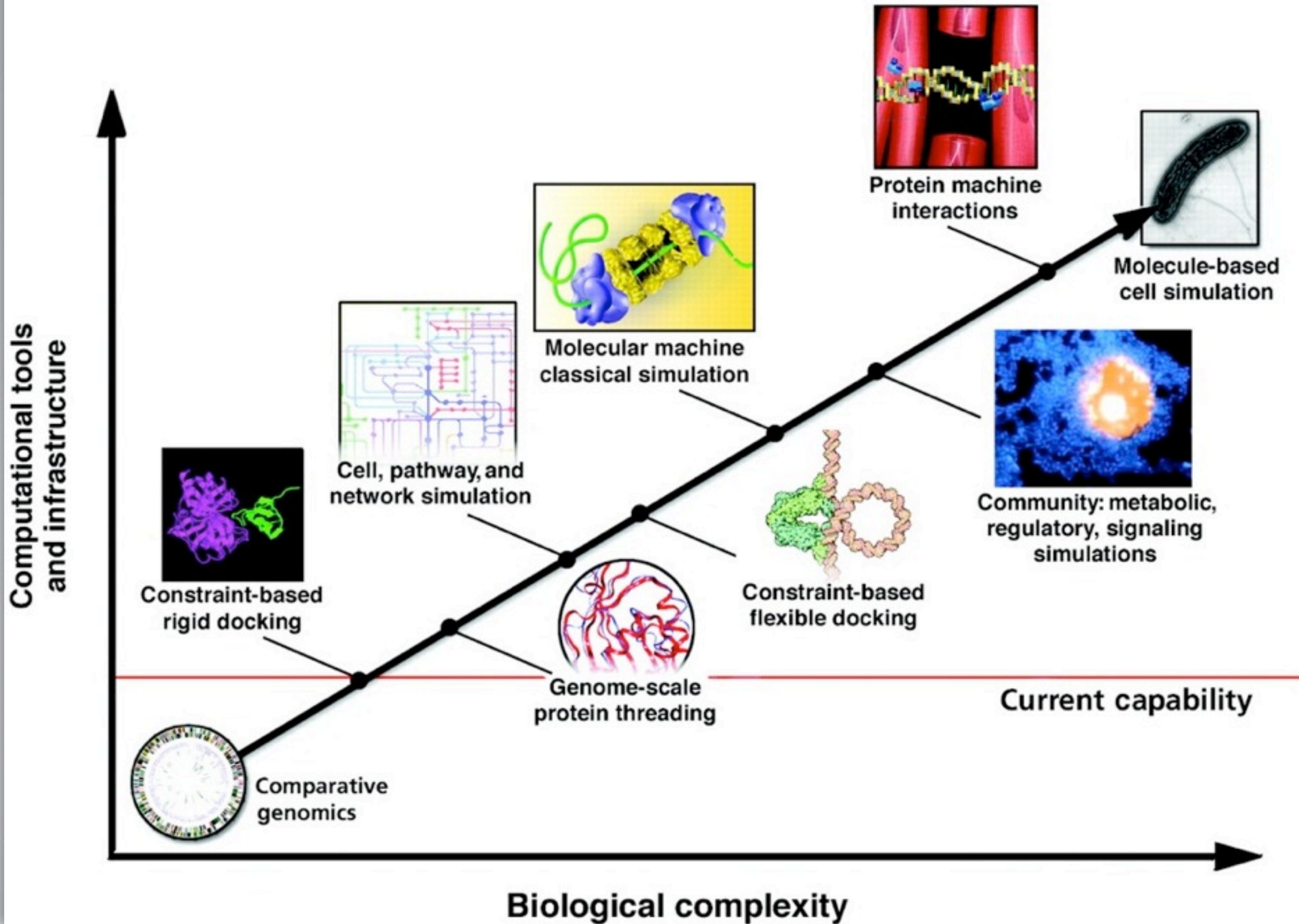


METAGENOMICS II

- Soil samples
- Sea water samples
- Air samples
- Medical samples
- Farm animal samples
- Ancient bones
- Human microbiome



Genomes to Life Computing



THE MESSAGE

- Everyone having a biological question can produce terabytes of sequence data at almost no cost
- Most groups run into storage and data analysis problems
- New kind of easy of use but safe Life Science Grid systems are needed for the future