



**Workshop: BioinfoGrid HealthGrid 2007 24th > 27th  
April 2007 Geneva, Switzerland**

# **BIOINFOGRID: BIOINFORMATICS SIMULATION AND MODELING BASED ON GRID**



**Milanesi Luciano**

National Research Council

Institute of Biomedical Technologies, Milan, Italy

[luciano.milanesi@itb.cnr.it](mailto:luciano.milanesi@itb.cnr.it)



# Related EU projects



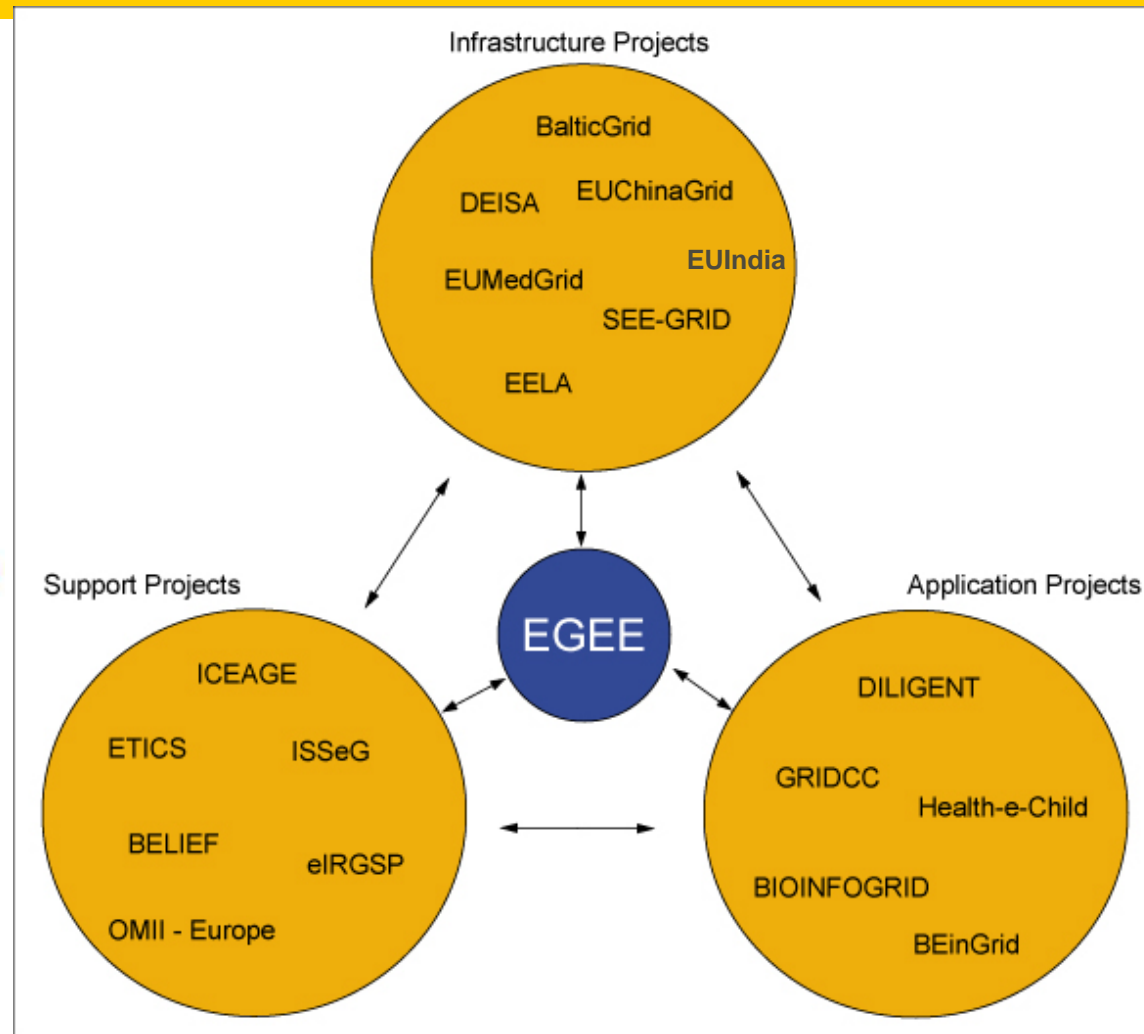
BioinfoGRID



ISSeG



eIRGSP





BioinfoGRID

# BioinfoGRID Project



- The **BIOINFOGRID project** proposes to combine the Bioinformatics services and applications for molecular biology users with the Grid Infrastructure by EGEE and EGEEII projects.
- In the BIOINFOGRID initiative we plan **to evaluate genomics, transcriptomics, proteomics and molecular dynamics applications studies based on GRID technology.**
- The project start date: 1st January 2006
- The project finish date: 31 December 2007



# Genomics applications in GRID

**Aim** : use of computational GRID to analyse molecular biological data at the genomic scale

## Description

- **GRID Bioinformatics User Tools**: unification of larger groups of bioinformatics tools into single analytical steps and their optimization for GRID
- **GRID analysis of cDNA data**: computer- aided functional annotation of cDNAs in order to optimize sensitivity and specificity



# Genomics applications in GRID

- **GRID analysis of genomic databases:** integration of precomputed data, gene identification, differentiation of pseudogenes, comparative genome analysis, etc.
- **Multiple alignments:** testing of new algorithms for computationally very demanding alignment procedures, optimization for GRID.

```
PRTC      TWFLVGLVSWG-EGCGLLHNYGVYTKVSRYLWDWIHGHIRDKEAPQKSWAP-----
FA10     TYFVTGIVSWG-EGCARKGKYGIIYTKVTAFLKWDIDRSMKTRGLPKAKSHAPEVITSSPLK
FA7      TWYLTGIVSWG-QGCATVGHFGVYTRVSQYIEWLQKLMRSE-----PRPGVLLRAPFP
THROMBIN RWYQMGIVSWG-EGCDRDGKYGFYTHVFRLLKKWIQKVIDQFGE-----
FA9      TSFLTGLIISWG-EECAMKGKYGIIYTKVSRYVNWIKKTKLT-----
KALLIKREIN MWRLVGITSWG-EGCARREQPGVYTKVAEYMDWILEKTQS SDGKAQM QS PA-----
FA11     VVHLYGITSWG-EGCAQRE R PGVYTNVVEYVDWILEKTQAV-----
TRYB1    TWLQAGVVSWG-EGCAQPNRPGIYTRVTYYLDWIHHYVPKKP-----
TRYB2    TWLQAGVVSWG-EGCAQPNRPGIYTRVTYYLDWIHHYVPKKP-----
TRYA     TWLQAGVVSWD-EGCAQPNRPGIYTRVTYYLDWIHHYVPKKP-----
KLKE     --QLQGLVSWG M ERCA L P G Y P G V Y T N L C K Y R S W I E E T M R D K-----
CTRL     TWVLI GIVSWG-TKN CNVRA PAVYTRVSKFSTWINQVIAYN-----
```



# Proteomics Applications in GRID

**Aim** : use of computational GRIDs to analysis molecular biological data in proteomics

## Description

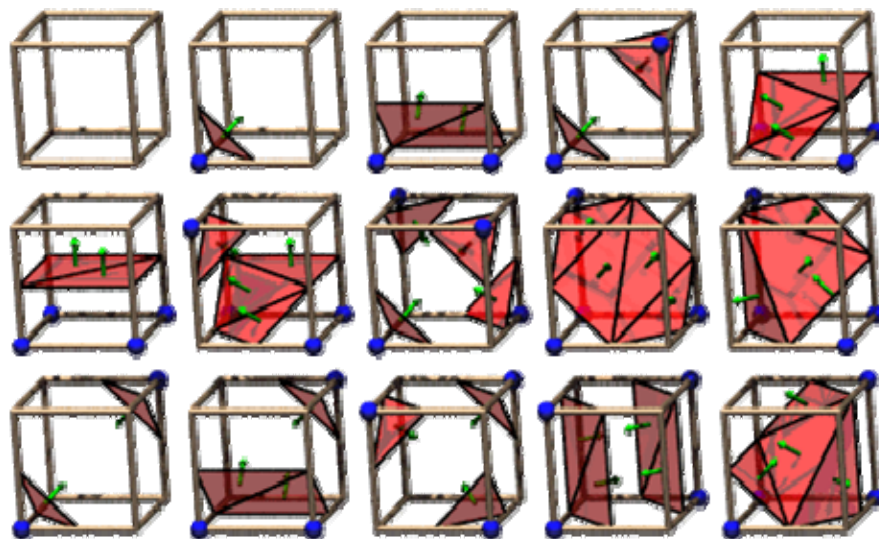
- **Perform functional protein analysis in GRID** by using the functional protein domain annotations on large protein families using GRID and related databases.



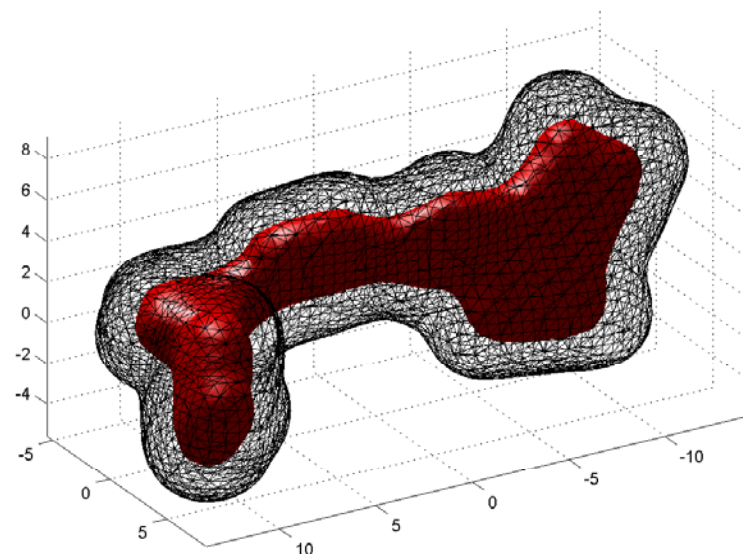


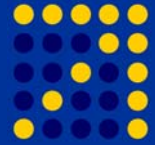
# Proteomics Applications in GRID

- **Protein surface calculation in GRID.** : the grid will be used to elaborate the volumetric description of the protein obtaining a precise representation of the corresponding surface.



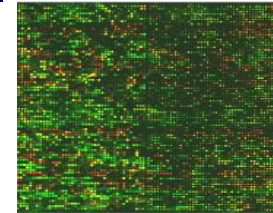
The 15 Cube Combinations





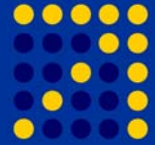
# Transcriptomics applications

**Aim** : use of computational GRIDs to analyse transcriptomics data and to perform application of Phylogenetic methods based on estimates trees.



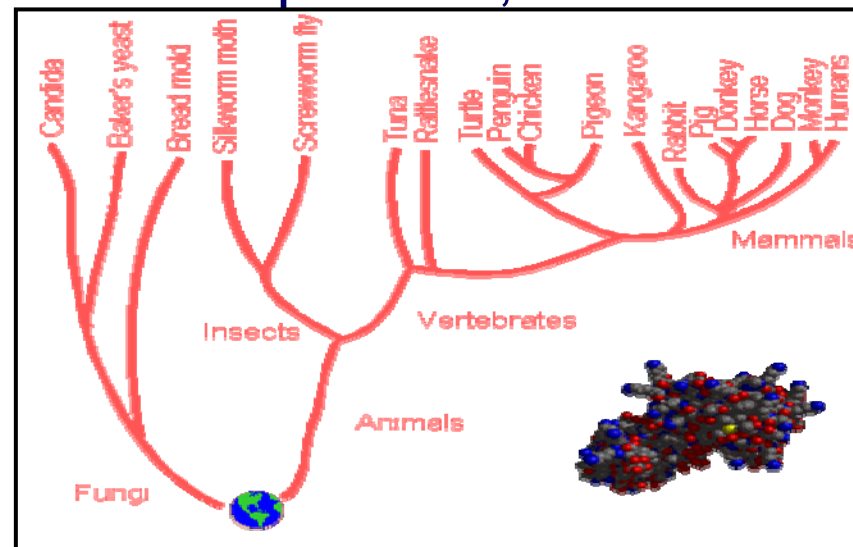
## Description

- **To perform algorithmic tools for gene expression data analysis in GRID:** evaluate the computational tools for extracting biologically significant information from gene expression data.
- Algorithms will focus on clustering steady state and time series gene expression data, multiple testing and meta analysis of different microarray experiments from different groups, and identification of transcription sites.



# Phylogenetic application in GRID

- **Phylogenetics** : Reconstructing the evolutionary history of a group of taxa is major research thrust in computational biology and a standard part of exploratory sequence analysis. An evolutionary history not only gives relationships among taxa, but also an important tool for inferring the universal **tree of life**, inferring structural, physiological, and biochemical properties of sequences from other similar sequences, and reconstruction of tissue evolution.





# Database Applications in GRID

**Aim :** To manage the biological database, by using the GRID infrastructure.

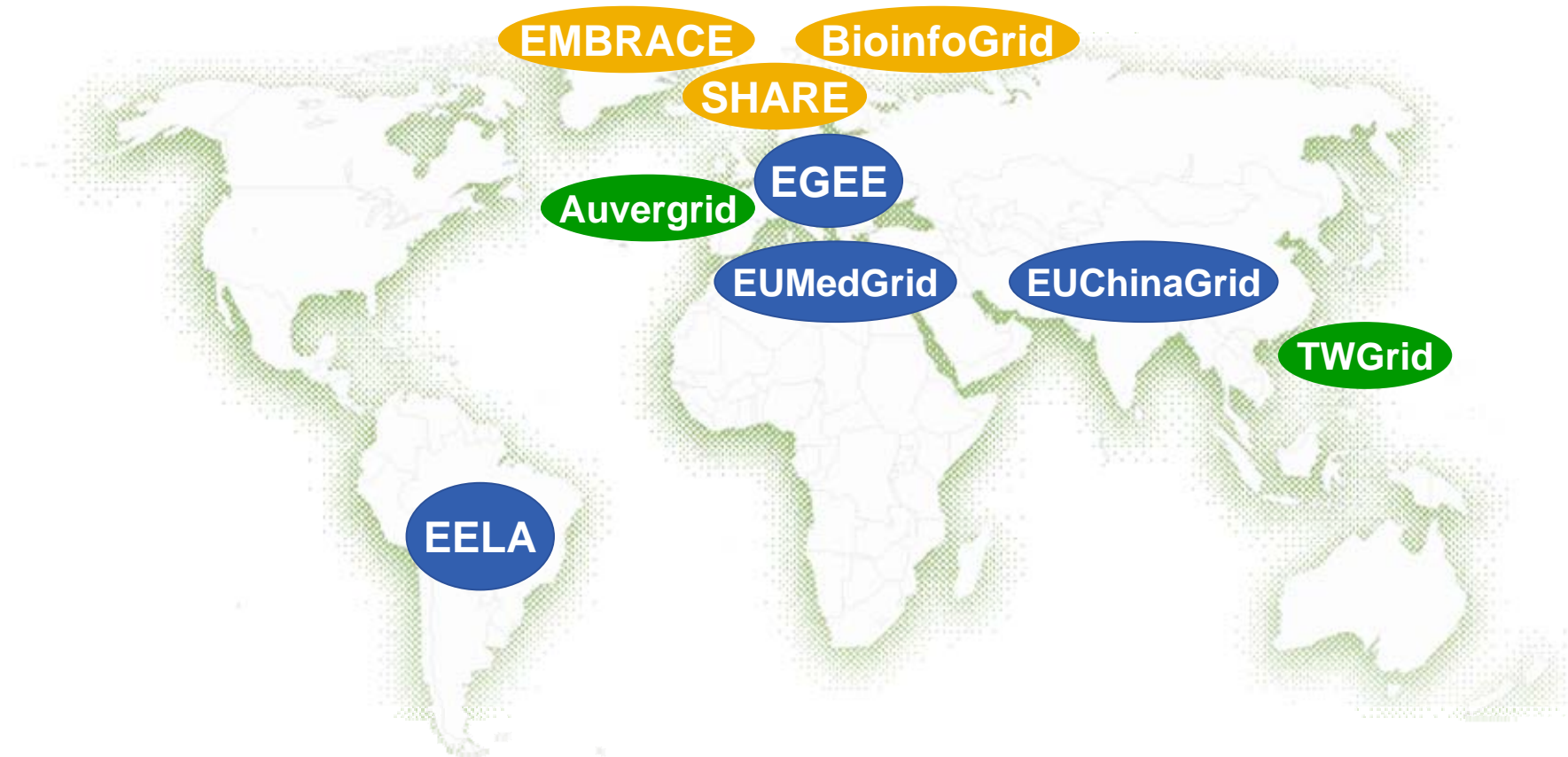
## Description

- **Biological database on GRID:** these databases will be complemented by others that are publicly available in Internet, by using GRID and web services where appropriate.
- **Functional Analogous Finder:** By using the GO terms and the associations to gene products it is possible to compare the total associated GO terms and their ascending parents to validate the functional analogy between two gene products





# Networks of people

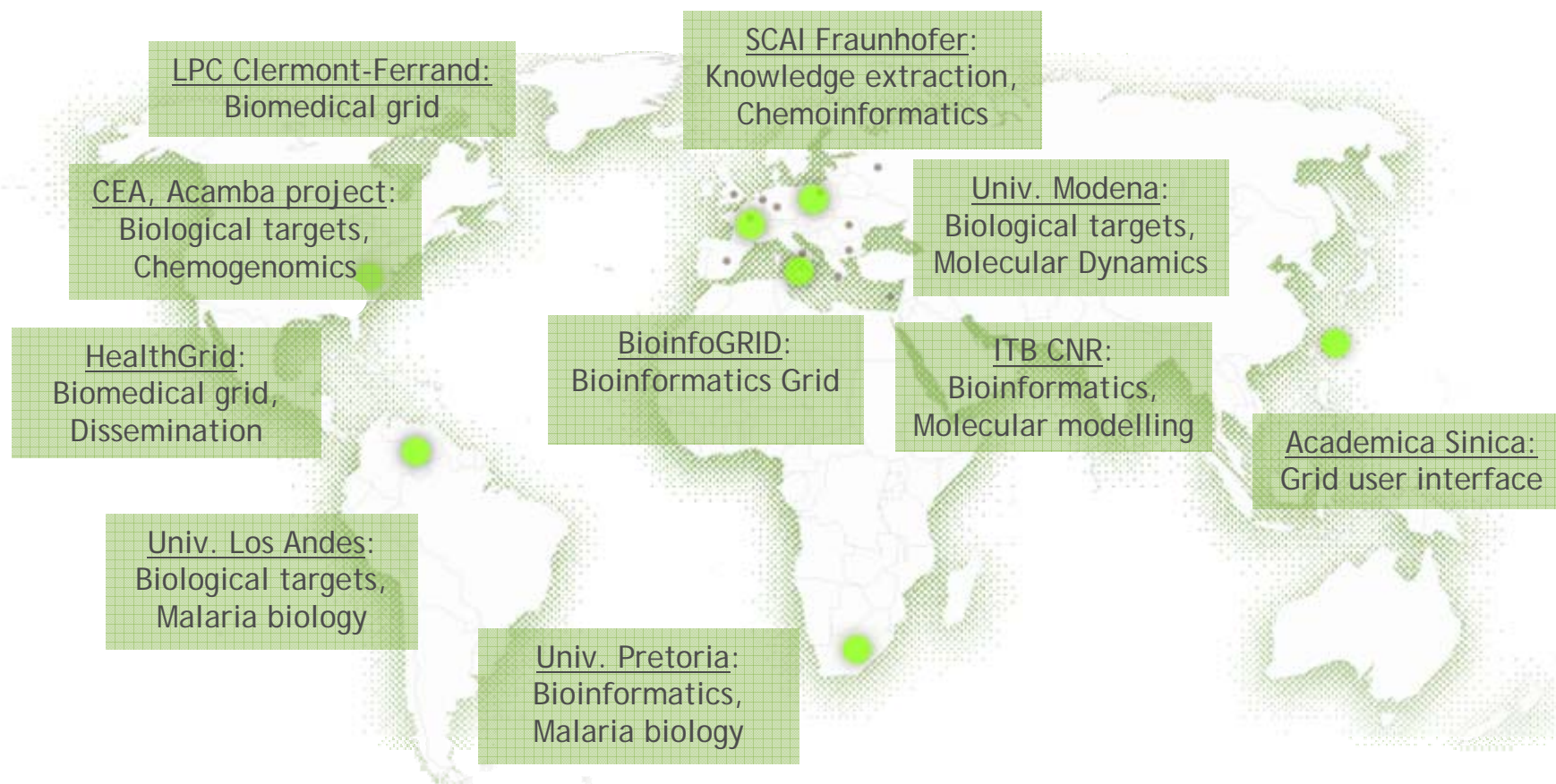


- : European grid infrastructure
- : European grid project
- : Regional/national grid infrastructure



BioinfoGRID

# A grid for Malaria and Avian Influenza



Use the grid technology to foster research and development on malaria and other neglected diseases

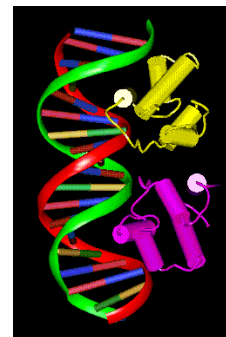
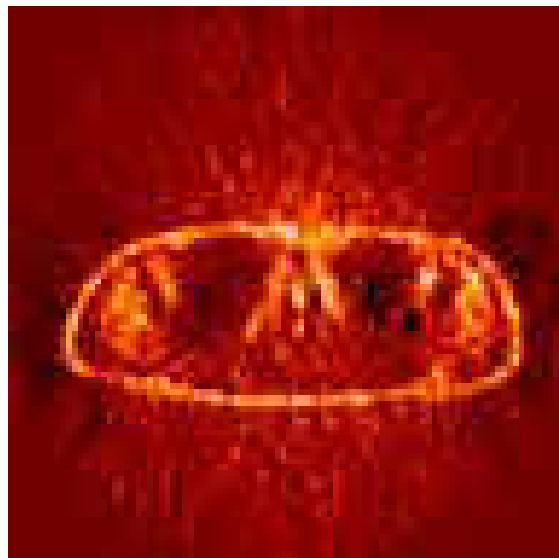
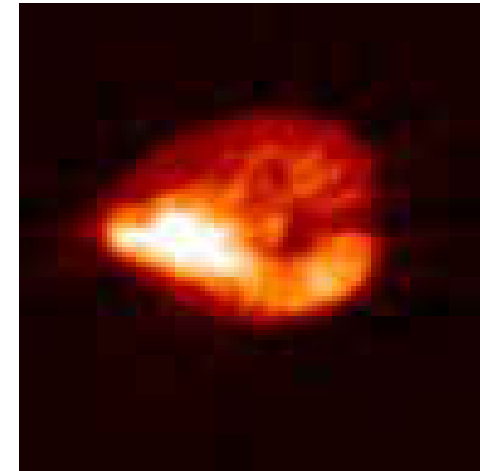
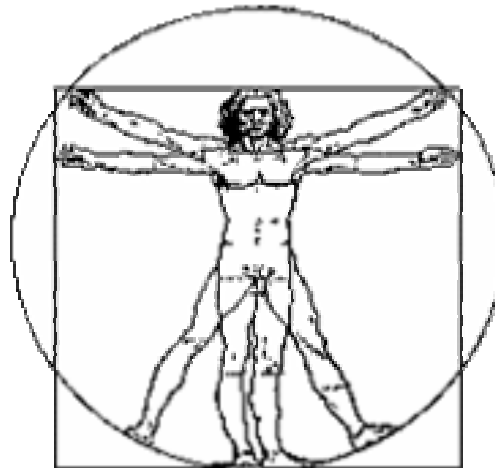
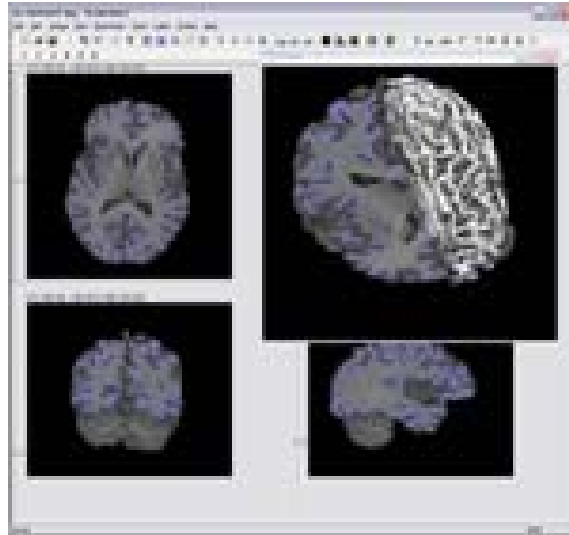
Contacts also established with WHO, Microsoft, TATRC, Argonne, SDSC, SERONO, NOVARTIS, Sanofi-Aventis, Hospitals in subsaharian Africa,

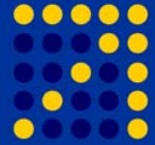




# System Biology for Health

BioinfoGRID



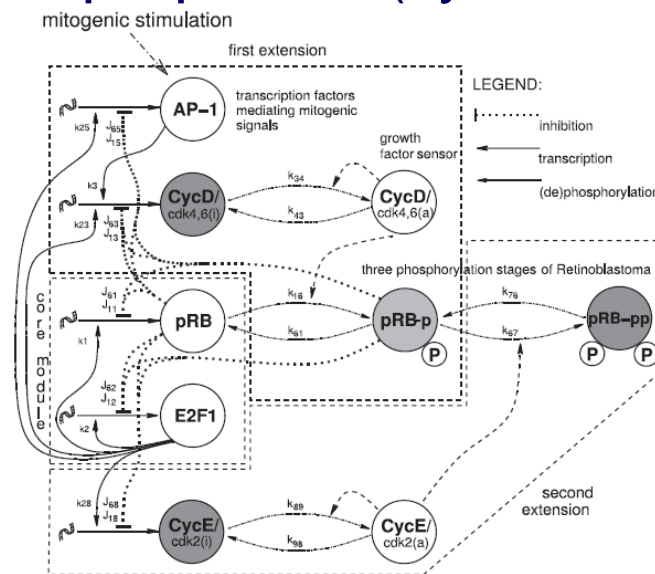


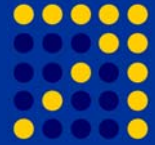
- Systems biology studies **how biological functions emerge** from the protein-protein interactions in the living systems;
- The **complexity** of this biological process relies in the high number of genes and networks of protein interactions involved in;
- The **quantification** of the behavior of each cell cycle components has a crucial role in the understanding the complex mechanism of cell cycle regulation.



- **What is modeling?**

- The act of representing something, usually on a smaller scale (general definition)
- Design and analysis of a mathematical representation of a biological system to outline unknown properties of the system, the emergent properties (systems biology definition)





- Simulation of an **ODE system** is possible on a single workstation: the numerical integration of an ODE system is not very time consuming
- **Parameter estimation**, the evaluation of the best set of parameters which define the model relating to a specific experimental dataset, **requires High Performance Computing techniques** since the complexity of finding the best solution to modelling.
- The estimation of the kinetic parameters *in silico* is performed by **fitting the data by computing a number of ODE systems** with different parameters and verifying the best solution.



BioinfoGRID

# Cell Cycle Database Model Section



**CCDB** Cell Cycle Database

- Home page
- Gene search
- Protein search
- Text search
- BLAST search
- Models
- Links
- Acknowledgements

[Publication paper](#)
[SBML components - formulas](#)
[Simulate this model](#)

## Bifurcation analysis of the regulatory modules of the mammalian G1/S transition.

(\*) Swat M, Kel A, Herzel H - 2004 - *Bioinformatics*

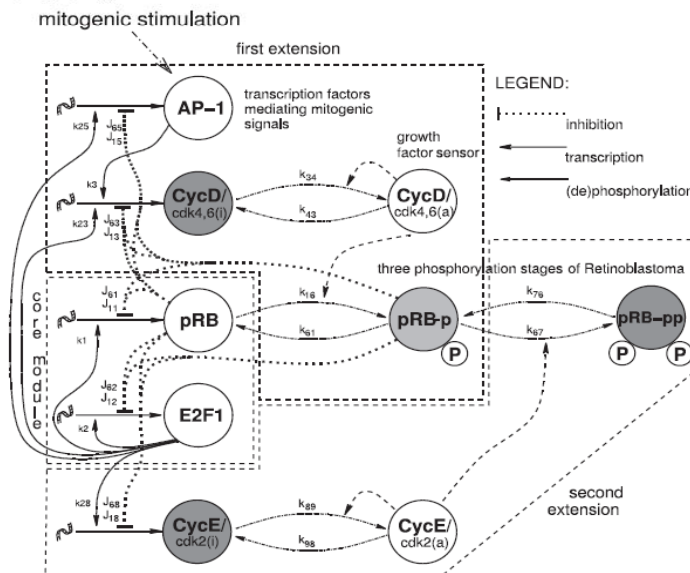
### Abstract:

MOTIVATION: Mathematical models of the cell cycle can contribute to an understanding of its basic mechanisms. Modern simulation tools make the analysis of key components and their interactions very effective. This paper focuses on the role of small modules and feedbacks in the gene-protein network governing the G1/S transition in mammalian cells. Mutations in this network may lead to uncontrolled cell proliferation. Bifurcation analysis helps to identify the key components of this extremely complex interaction network. RESULTS: We identify various positive and negative feedback loops in the network controlling the G1/S transition. It is shown that the positive feedback regulation of E2F1 and a double activator-inhibitor module can lead to bistability. Extensions of the core module preserve the essential features such as bistability. The complete model exhibits a transcritical bifurcation in addition to bistability. We relate these bifurcations to the cell cycle checkpoint and the G1/S phase transition point. Thus, core modules can explain major features of the complex G1/S network and have a robust decision taking function.

### Organism:

Mammalian

### Paper graph (\*):



### Model proteins:

- CCND1
- RB
- E2F1
- CDK4
- CDK6
- CDK2
- CCNE2
- CCNE1
- JUN\_HUMAN
- CCND2
- TDP1
- TDP2

### Links

- PubMed entry





## Reactions

Reaction name (id if name not defined)	reversible	reactants	direction	products	kinetic law
pRB synthesis	false	-	=>	pRB	$k_1 \frac{E2F1}{Km1 + E2F1} \frac{J11}{J11 + pRB} \frac{J61}{J61 + pRBp}$
pRB phosphorylation	false	pRB	=>	pRBp	$k_{16} pRB \text{ CycDa}$
pRBp dephosphorylation	false	pRBp	=>	pRB	$Mass\_Action\_1(k_{61}, pRBp)$
pRB degradation	false	pRB	=>	-	$Mass\_Action\_1(\phi_{pRB}, pRB)$
E2F1 synthesis	false	-	=>	E2F1	$k_{p+} \frac{k_2(a^2 + E2F1^2)}{Km2^2 + E2F1^2} \frac{J12}{J12 + pRB} \frac{J62}{J62 + pRBp}$
E2F1 degradation	false	E2F1	=>	-	$Mass\_Action\_1(\phi_{E2F1}, E2F1)$

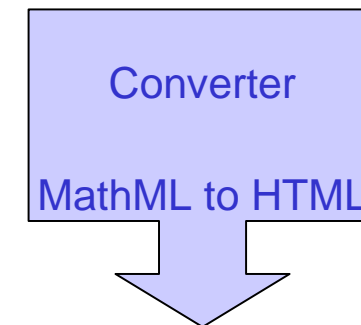
## Species

id	name	compartment	initial amounts	initial concentration	unit
pRB_1	pRB	cell_1	0.1	-	mole
pRBp_1	pRBp	cell_1	0.1	-	mole
E2F1_1	E2F1	cell_1	0.1	-	mole
CycDi_1	CycDi	cell_1	0.1	-	mole
CycDa_1	CycDa	cell_1	0.1	-	mole
AP1_1	AP1	cell_1	0.1	-	mole
pRBpp_1	pRBpp	cell_1	0.1	-	mole
CycEi_1	CycEi	cell_1	0.1	-	mole
CycEa_1	CycEa	cell_1	0.1	-	mole

## Ordinary Differential Equations

left side	right side
$\frac{d pRB}{dt} =$	$k_1 \frac{E2F1}{Km1 + E2F1} \frac{J11}{J11 + pRB} \frac{J61}{J61 + pRBp} + Mass\_Action\_1(k_{61}, pRBp) - (k_{16} pRB \text{ CycDa}) - Mass\_Action\_1(\phi_{pRB}, pRB)$
$\frac{d pRBp}{dt} =$	$k_{16} pRB \text{ CycDa} + Mass\_Action\_1(k_{76}, pRBp) - Mass\_Action\_1(k_{61}, pRBp) - (k_{67} pRBp \text{ E2F1}) - Mass\_Action\_1(\phi_{pRBp}, pRBp)$
$\frac{d E2F1}{dt} =$	$k_{p+} \frac{k_2(a^2 + E2F1^2)}{Km2^2 + E2F1^2} \frac{J12}{J12 + pRB} \frac{J62}{J62 + pRBp} - Mass\_Action\_1(\phi_{E2F1}, E2F1)$
$\frac{d CycDi}{dt} =$	$k_3 AP1 + k_{23} E2F1 \frac{J13}{J13 + pRB} \frac{J63}{J63 + pRBp} + Mass\_Action\_1(k_{43}, CycDa) - (k_{34} CycDi \frac{CycDa}{Km4 + CycDa}) - Mass\_Action\_1(\phi_{CycDi}, CycDi)$
$\frac{d CycDa}{dt} =$	$k_{34} CycDi \frac{CycDa}{Km4 + CycDa} - Mass\_Action\_1(k_{43}, CycDa) - Mass\_Action\_1(\phi_{CycDa}, CycDa)$
$\frac{d AP1}{dt} =$	$F_m + k_{25} E2F1 \frac{J15}{J15 + pRB} \frac{J65}{J65 + pRBp} - Mass\_Action\_1(\phi_{AP1}, AP1)$

## SBML model



## SBML visualization



## Species\*

Species id	Value	Species id	Value	Species id	Value	Species id	Value	Species id	Value					
pRB_1	0.1	mole	pRBp_1	0.1	mole	E2F1_1	0.1	mole	CycDi_1	0.1	mole	CycDa_1	0.1	mole
AP1_1	0.1	mole	pRBpp_1	0.1	mole	CycEi_1	0.1	mole	CycEa_1	0.1	mole			

## Parameters

Parameter id	Value	Parameter id	Value	Parameter id	Value	Parameter id	Value	Parameter id	Value
k1_1	1.0	Km1_1	0.5	J11_1	0.5	J61_1	5.0	k16_1	0.4
k61_1	0.3	phi_pRB_1	0.0050	kp_1	0.05	k2_1	1.6	a_1	0.04
Km2_1	4.0	J12_1	5.0	J62_1	8.0	phi_E2F1_1	0.1	k3_1	0.05
k23_1	0.3	J13_1	0.0020	J63_1	2.0	k34_1	0.04	Km4_1	0.3
phi_CycDi_1	0.023	k43_1	0.01	phi_CycDa_1	0.03	Fm_1	0.0050	k25_1	0.9
J15_1	0.0010	J65_1	6.0	phi_AP1_1	0.01	k67_1	0.7	k76_1	0.1
phi_pRBpp_1	0.04	phi_pRBp_1	0.06	k28_1	0.06	J18_1	0.6	J68_1	7.0
k89_1	0.07	Km9_1	0.0050	k98_1	0.0010	phi_CycEi_1	0.06	phi_CycEa_1	0.05

## XPPAUT internal options setting

total	dt	method	tolerance	minimum step	maximum step	delay
500	0.1	stiff	0.001	1e-12	1	0

\* when not provided in the original paper we suggest possible values

The simulation software is XPPAUT - Bard Ermentrout, 2002

- dt = sets step size used by the fixed step integrators and the output step for Gear and Stiff
- total = sets total time
- tolerance = sets the error tolerance for adaptive methods
- sets the minimum allowable time step for adaptive methods
- sets the maximum allowable time step for adaptive methods
- delay = sets the upper bound for the maximum delay, for use with Delay Differential Equation

Users can:

➤ Change parameter values and protein

initial concentration

➤ Select the ODE solver



## Simulation results

Swat M, Kel A, Herzel H: Bifurcation analysis of the regulatory modules of the mammalian G1/S transition. - 2004

Download XPPAUT input file

Download results file\*

### Select species to show on 2D plot

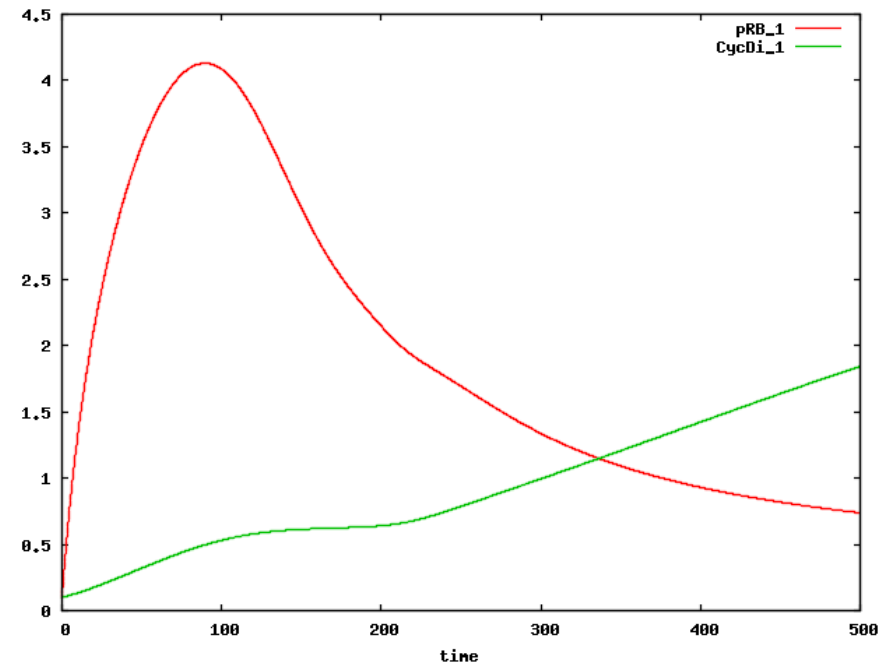
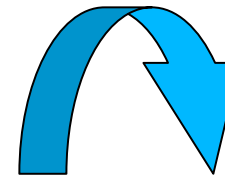
**x**  
time

**y series**

-	-	-	-	-
-	-	-	-	-

time  
pRB\_1  
pRBp\_1  
E2F1\_1  
CycDi\_1  
CycDa\_1  
AP1\_1  
pRBpp\_1  
CycEi\_1  
CycEa\_1

with GNUPLOT  
sets order of variables in the input interface; first is time



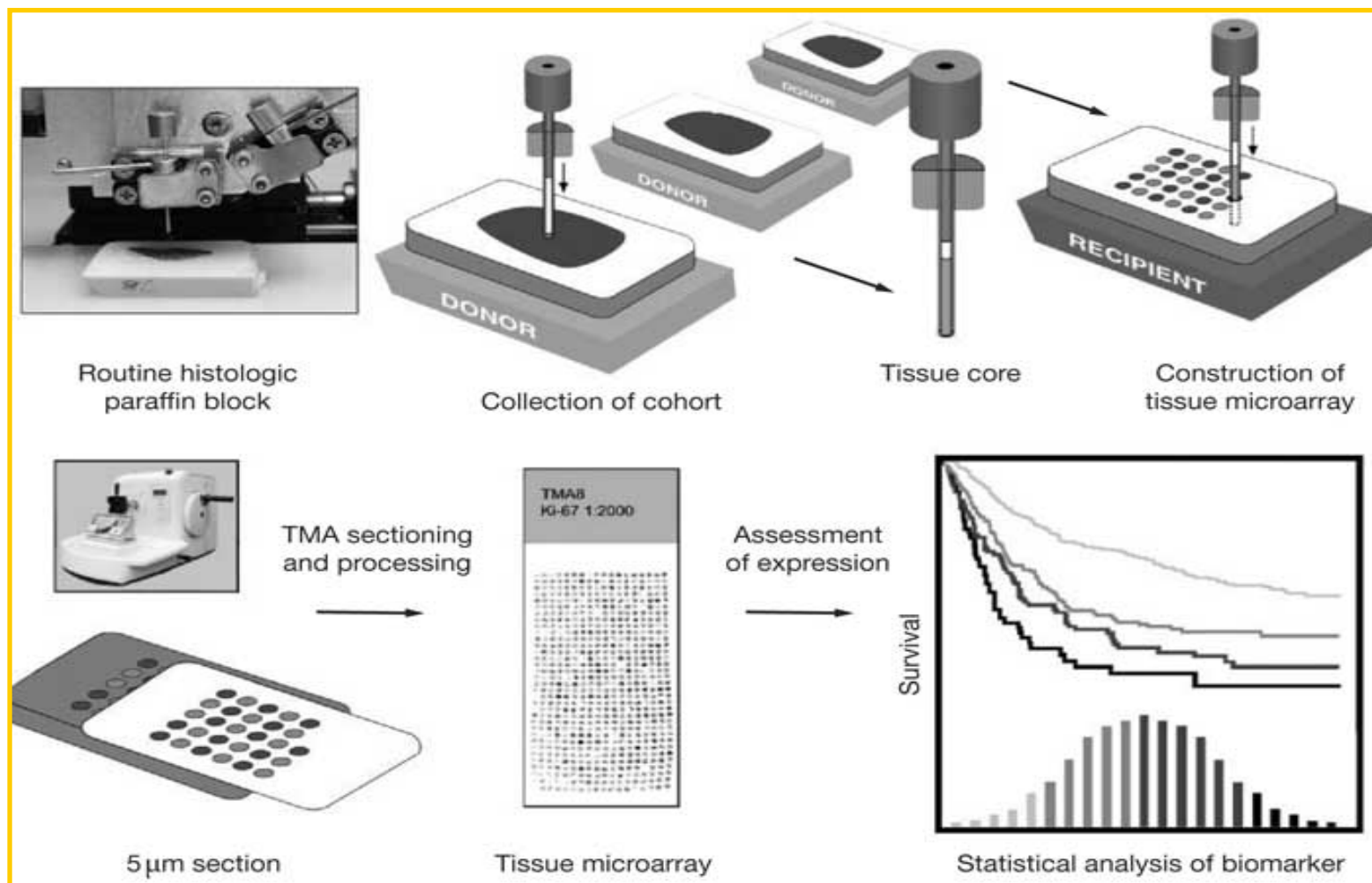
2D plot: image exported in png using GnuPlot

The simulation of a single ODE system describing a cell cycle model is possible



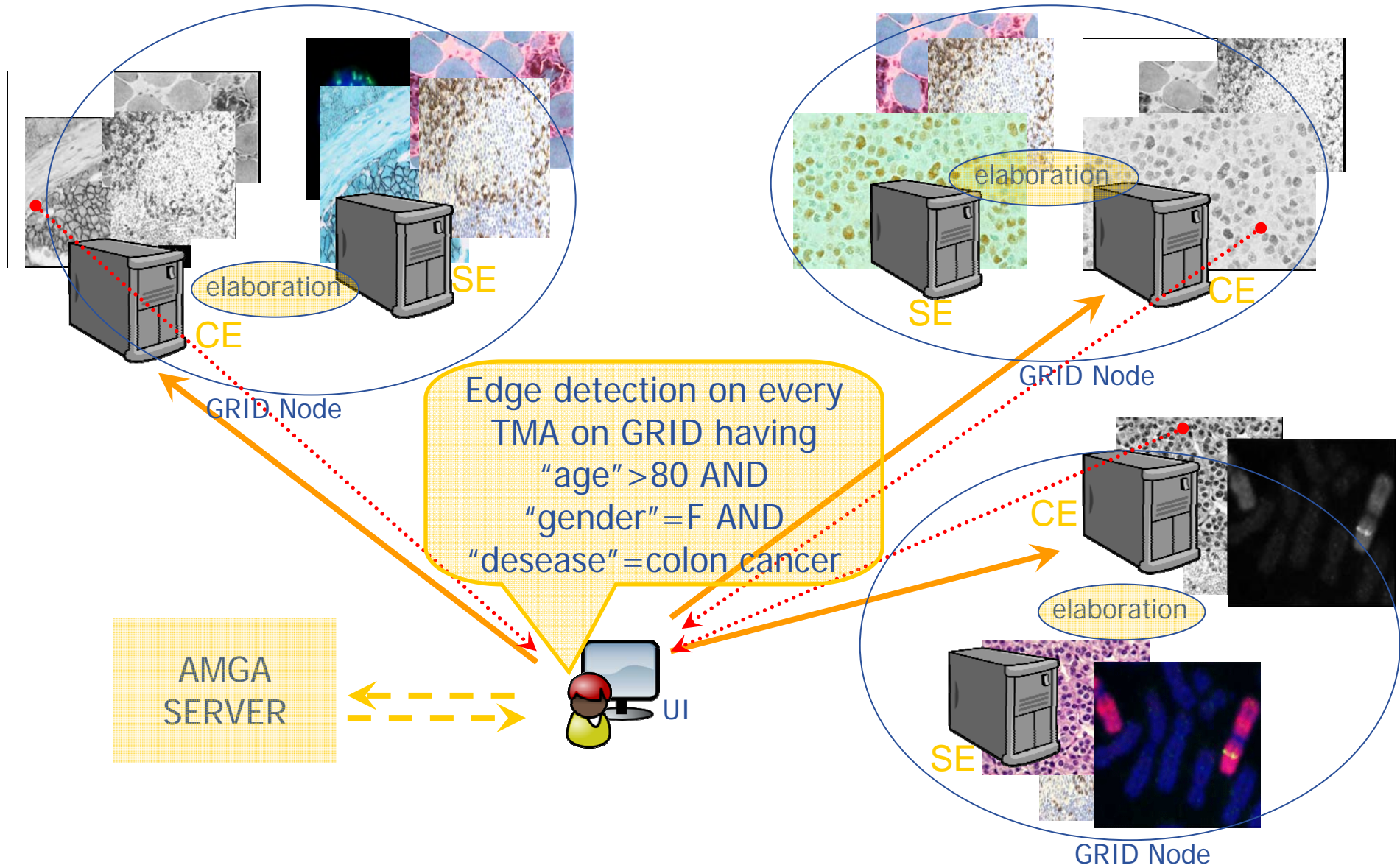
# Tissue Microarray technology

## Genes and proteins detection





# TMA on GRID





BioinfoGRID

# Acknowledgments



- BioinfoGRID  
<http://www.bioinfoGRID.eu>



- EGEE Enabling Grid for E-science project  
<http://www.eu.egEE.org>



- FIRB-MIUR LITBIO: Laboratory for Interdisciplinary Technologies in Bioinformatics  
<http://www.litbio.org>,