Project title: Open Screening Environment

Open and module-oriented system as a new platform that supports high content screening (HCS) in biology and medicine[1]

Project web page: http://sen.sourceforge.net

Abstract

In the post-genomic era, large-scale screens have become more and more common. This has led to the emergence of a new field of research coined "systems biology", which relies on the analysis of biological processes and pathways with large-scale experiments. Such experiments produce large data sets that require unbiased and automated analysis. Our system is meant to manage and analyze these data sets.

Here we present an idea of an ICT platform that supports high-throughput research in modern biology and medicine. The project aim is:

1) to gather research tools, modules that manage, analyze and process data sets generated during biological experiments,
2) to combine them into a one dynamic, interoperable, practical application to be used by different research actors involved in HCS.

Modules of interest perform specific but independent tasks from research fields related to HCS such as bioinformatics, data mining, structural genomics, image analysis, visualization, phenotyping and genotyping. By communicating with each other, the complementary applications would allow to track a complete screening process. Such approach provides a front-end solution to institutions, companies and universities that perform high-throughput research.

Motivation

There exists a number of commercial systems that analyze HCS data. However, they are very expensive, provide a limited insight into the algorithmic part and very rarely offer a possibility to incorporate own solutions. Although they contain sophisticated processing and analysis tools, the analysis of customized research tasks that are not addressed by commercial systems is restricted or cost extra money.
Moreover, it is often hard to anticipate which of commercial software packages would suffice to analyze data in future experiments.

i) OSE as an alternative to commercial systems by providing a similar or better functionality at lower cost.

In the research community there is a wide range of open-source applications that solve various tasks related to HCS. However, the combination of results generated by these application is restricted, as they do not provide a common interface. The results generated from the analysis of images does not often meet standards of data mining software. A large number of data generated during experiments are analyzed from from different point of view. The combination of results is difficult as there is no universal data exchange standard. Therefore, OSE is also meant as a standard of data exchange format.

---

[1] Prepared by Dept. of Automation and Modeling at Wroclaw University of Technology.
Created on 13th of Jan 2007, updated on 18th of Jun 2007.

ii) OSE as a platform of independent but complementary software and as a data exchange format

To our knowledge there exists no open source solution that supports high throughput biological screenings with software capable to manage and analyze data generated during experiments. Our existing and future software solutions will fill this gap.
The systematic approach could speed up and enhance the recognition of function of genes. It could increase knowledge of biological processes and mechanisms involved in diseases like cancer or Alzheimer and could have an impact on the development of new gene therapies and therapeutic interventions.

Current objectives

1. **Application to EU Framework Programs** (FP7 Health, deadline on 18th of Sep 2007 or FP7 ICT deadline on 9th of Oct 2007)
    1. Finding a coordination institution.
       An experienced coordinator would have to:
       1) prepare the project proposal (part A (information about participants) and B (project description, staff effort, milestone, deliverable list, sub-contracting, etc.) according to FP7 requirements. It means managing and gathering the description from each of partners and form a final project proposal
       2) plan working pachages, project timeline and the budget
       3) coordinate the flow of information between partners,
       4) split the tasks between partners and keep them on schedule
       5) consult with partners the FP7 call (this should be done asap)
       6) coordinate the project once it is approved by the EU Comission
       7) contact the EU Commision on behalf of the consortium
    2. Finding a matching FP7 Call of proposal (see a list of potential calls at the end of the document)
2. **Improvement and adaptation of already developed modules.**

   First, we need to modify the architecture of each of the modules to combine them into one, extensible and coherent system. We want to allow inter-modules communication and, at the same time, leave the possibility for modules to work independently. We want the system to be easily extensible with new algorithms and new functionality, so that at the later time we mostly focus on algorithms development and less on pure programming.
3. **Definition of the data exchange format** (one option would be to adopt Open Microscopy Environment or develop a XML-based format)
4. **Setting up research goals in other fields related to gene function study.**
   We plan to find and get in touch with other computer processing groups in EU and other countires that are doing statistical pattern recognition, data mining, image processing. The next step would be to develop new algorithms to analyze data provided by partners.
5. **Extending the collaboration with other partners from EU and other countries.**
   We want to find groups in research institutions that would benefit from using the OSE and present software to them. We also plan to participate in conferences to present the OSE and build new contacts. The other part of the activity would be supervising Ph.Ds, diploma thesis and student projects that are related to the OSE project.

The present state of the OSE

The OSE is an Open Source application. It will contribute to better development and distribution of the system. Furthermore, the platform where a system is developed and database engine are free and available for many operating systems (Mac OS, Windows, Linux) which make them available to the research community and reduce costs associated with development and usage.

The system will be designed for managing and analyzing all areas of high throughput screening experiments. It will integrate information about produced biological compounds, their use (cell line, plate and well position), screen parameters (date, experiment conditions, equipment) and imaging techniques applied. The system will track the complete screening process starting from production of biological and chemical compounds, their application in assays and will evaluate their influence on cells by analyzing images generated during experiments.

A number of modules are already deployed. The image processing module (Biomage) is able to store and process images or movies. It generates feature vectors that can be stored, evaluated and used to extract phenotypes with appropriate database queries. Thus, the module enables searching for compounds having specific influence on cells - specific phenotypes. The screen results can be public available via a user-friendly interface facilitating post-processing steps. The TracePilot module [1] visualizes results in 2D or 3D and performs analysis in terms of recognition groups of cells that show similar features (behavioral, dynamic or other). Systems' plug-in architecture simplifies the addition of new algorithms for data transformation and analysis of images and data.

Currently, OSE system consists of the following ready-to-use modules:

1 BioImage: image processing, detection of phenotypes

Processes images or movies and detects phenotypes. Simplifies the addition of dedicated image processing algorithms that solve various tasks. Maintained by MPI-CBG. The software will be available in August.

2. BioImageXD: image processing

Free open source software for analysis, processing and 3D rendering of multi dimensional microscopy images. Designed and developed by microscopists, cell biologists and programmers from the Universities of Jyväskylä and Turku in Finland, and collaborators worldwide.

3. TracePilot: object visualization

Visualizes objects over time in 2D or 3D. Allows discovery of common objects features. Published and maintained by BIOTEC.

4. KNIME: Konstanz Information Miner

Data mining and statistical pattern recognition application. Maintained and published by Department of Applied Computer Science Bioinformatics and Information Mining in Konstanz University.

5. Dynomica Epidetect

Models the immune response by predicting and designing a possible B-cell epitopes from the protein sequence. Developed by Dynomica, Ltd.

6. Cell Detective

3D cell tracking in confocal microscopy. Developed for MPI-CBG. Currently under development.

Potential partners

* Max Planck Institute of Cell Biology and Genetics, Germany (generation of experimental data, screenings development, High Content Screening, BioImage)
* Konstanz University, Department of Computer and Information Science, Germany (data mining, KNIME)
* BioImageXD, Finland (image processing, BioImageXD)
* Wroclaw University of Technology, Poland (data mining, hardware acceleration)
* BIOTEC, Germany (bioinformatics, molecular docking, TracePilot, BioImage)
* Wroclaw Technology Park, Poland (laboratory facility)
* Vratis (SME, image processing, pattern recognition, software development)
* Dynomica, Bulgaria (SME, Bioinformatics, software development)
* Univeristy of Applied Sciences in Heigelberg, Austria (bioinformatics),
* Linz University, Austria (data mining)

Possible FP7 calls

FP7 ICT-2007.5.3: Virtual Physiological Human

Data integration and new knowledge extraction: Innovative software tools for data mining, representation, formalisation and image processing able to integrate heterogeneous multimedia information from distributed databases. These tools will be developed specifically for (1) Coupling scientific research data with clinical and large empirical databases with focus on the association of genotype-related data and phenotype-related data with specific computational models of diseases and treatments;

(2) Automated image processing and analysis for the extraction of bio-medical parameters/markers used to assess the presence or evolution of a disease, focusing on specific organs and/or disease and demonstrating quantitative benefits in diagnosis and prognosis. Applications: (1) Prediction of disease or early diagnosis by integrating patient specific knowledge and predispositions obtained in biomedical imaging; (2) Advanced environment for simulation and assessment of the efficacy and safety of specific drugs.

FP7 HEALTH-2007-1.1-4 SME-driven collaborative research projects for developing tools and technologies for high-throughput research.

The focus should be on developing and improving tools and technologies for: cell based assays, sequencing; gene expression, proteomics, genotyping and phenotyping; structural genomics; bioinformatics and systems biology; metabolomics, other “omics”. These SME-driven projects should be specifically designed to encourage industry, preferably SME efforts towards research and innovation. The expected project results should clearly be of interest and potential benefit to SME(s). All consortia should have at least 40 % of the requested EC contribution budget going to SMEs. Funding scheme: Collaborative projects (Small or medium-scale focused research projects targeted to SMEs).

FP7 HEALTH-2007-2.1.2-5: Multidisciplinary fundamental genomics and molecular biology approaches to study basic biological processes relevant to health and diseases.

Projects should be multidisciplinary and should focus on collecting, analysingand applying quantitative data

to enable system biological approaches addressing basic biological processes at all appropriate levels of system complexity.

FP7 HEALTH-2007-1.4-4: Development of emerging gene therapy tools and technologies for clinical application.

This project should aim to exploit emerging gene therapy tools and technologies, such as use of genome editing and repair (RNAi, site specific recombination, etc.) to correct genetic defects, or new gene transfer techniques (novel virus vectors, targeted nanoparticles, bacterial DNA-based vector, bactofection, etc.), which overcome the limitations of existing tools and for which preclinical proof of concept has already been established. The project should address biological activity, pharmacokinetics and toxicology of the gene therapy vector, establish biomarkers and assays for the evaluation of clinical trials, and take a translational approach towards early clinical research for therapeutic intervention. It should mobilise industrial and academic partners and address ethical and regulatory issues.

FP7 HEALTH-2007-1.4-5: Gene therapy tools targeting the central nervous system.

Projects should aim to meet specific challenges posed by the central nervous system for gene therapy. Research should focus on developing and validating new gene therapy tools inducing long-lived, safe, cell-type specific and regulated transgene expression in the central nervous system for application in therapy of neurological disorders. Funding scheme: Collaborative projects (Small or medium-scale focused research projects).