CCO: a Cell Cycle Ontology

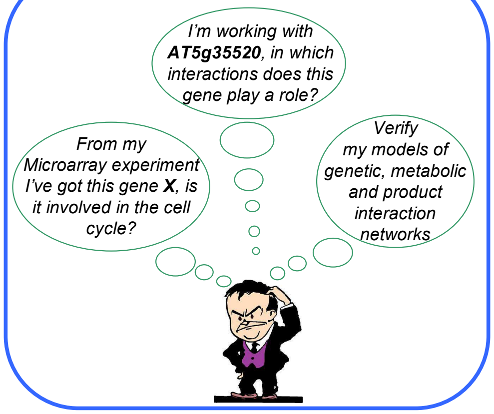
Erick Antezana, Mikel Egaña, Vladimir Mironov, and Martin Kuiper

Dept. of Plant Systems Biology. Flanders Interuniversity Institute for Biotechnology/Ghent University.

Technologiepark 927, B-9052 Ghent BELGIUM. {erant, miega, vlmir, makui}@psb.ugent.be

http://www.CellCycleOntology.org

Motivating questions



I'm working with ***AT5g35520****, in which interactions does this gene play a role?*

From my Microarray experiment I've got this gene ***X****, is it involved in the cell cycle?*

Verify my models of genetic, metabolic and product interaction networks

Prospective users

- **Molecular biologist:** interacting components, interacting events, and the roles that each component play. Hypothesis evaluation.
- **Bioinformatician:** data integration, annotation, modeling and simulation.
- **General audience:** education.

Background

The amount of data generated in biological experiments continues to grow exponentially. Current tools and approaches for analyzing this data avalanche only allow a limited extraction of information and knowledge. To make matters worse, the lack of a structured documentation format for this knowledge leaves much of the information extracted from the raw data unused. Differences, for instance, in the technical languages in use (**synonymy** and **polysemy**) have complicated the further interpretation of information.

Main Objective

To capture the semantics of the cell cycle regulatory process, especially the dynamic aspects of the concepts and their interrelations, and to promote sharing, reuse and enable better computational integration with existing resources.

Target organisms: *S. cerevisiae, S. pombe, A. thaliana and H. sapiens.*

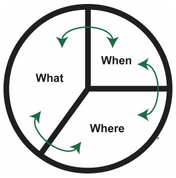


Fig 1. Combining cellular components, molecular functions and biological processes.

The elements combined by CCO are depicted in Fig. 1. A very simple knowledge sample: *« Cyclin B (what) is located in Cytoplasm (where) during Interphase (when) ».*

Tools

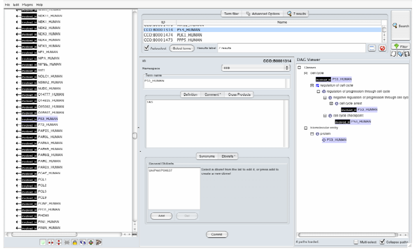


Fig 2. CCO using OBO Edit.

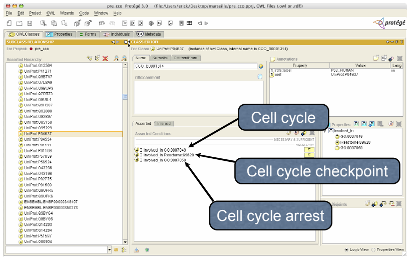
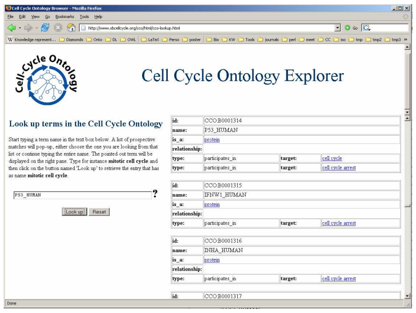


Fig 3. CCO using Protégé.



http://www.sbcellcycle.org/cco/html/cco-lookup.html

Cell Cycle Ontology Explorer

Look up terms in the Cell Cycle Ontology

Start typing a term name in the text box below. A list of prospective matches will pop-up, either choose the one you are looking from that list or continue typing the entire name. The pointed out term will be displayed on the right pane. Type for instance **mitotic cell cycle** and then click on the button named 'Look up' to retrieve the entry that has as name **mitotic cell cycle**.

P53_HUMAN ?

id:	CCO:B0001314		
name:	P53_HUMAN		
is_a:	protein		
relationship:			
type:	participates_in	**target:**	cell cycle
type:	participates_in	**target:**	cell cycle arrest

id:	CCO:B0001315		
name:	IFNW1_HUMAN		
is_a:	protein		
relationship:			
type:	participates_in	**target:**	cell cycle arrest

id:	CCO:B0001316		
name:	INHA_HUMAN		
is_a:	protein		
relationship:			
type:	participates_in	**target:**	cell cycle arrest

id:	CCO:B0001317

Fig 4. CCO Explorer (online).

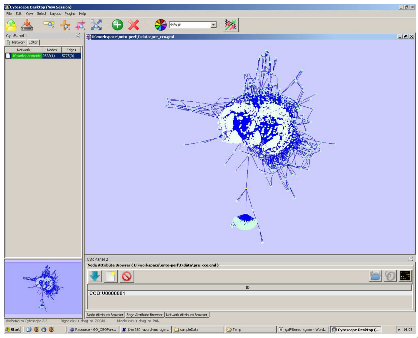


Fig 5. CCO in Cytoscape.

Method

CCO should capture the semantics, temporal aspects and dynamics of the cell-cycle process. CCO forms the knowledge base core. GO only considers subsumption (is_a) and partonomic inclusion (part_of).

Data sources:

- Association files.
- Interactions data: BIND, IntAct.
- Cell cycle functional data.
- Data obtained from bio-tools.

OWL-DL has been chosen for representing the knowledge. Used reasoners: RACER, KAON, Pellet and FaCT++.

Availability

Formats:

- OBO.
- OWL.
- XML.
- GML.
- XGMML (under development).
- SBML (under development).

For developers

- API in PERL for dealing with any OBO ontology like Gene Ontology (GO).
- API in JAVA (under development).
- SVN access (very soon).

Exploiting reasoning capability

There are a number of relationships that should have been annotated as *part_of* instead of *is_a* and vice versa.

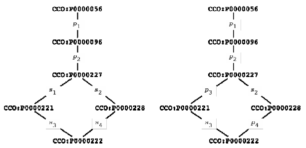


Fig 6. The sub-ontology on the left has inconsistent relation s4 (is_a) which has been changed into part_of (right side) .

CCO ID	Term
CCO:P0007049	cell cycle
CCO:P0000096	centrosome cycle
CCO:P0000227	regulation of centrosome cycle
CCO:P0000221	regulation of centriole replication
CCO:P0000228	negative regulation of centrosome cycle
CCO:P0000222	negative regulation of centriole replication

Conclusions and future work

- Developed: data integration pipeline prototype (the entire life cycle of the knowledge base).
- Concrete problems and initial results related to the implementation of automatic format mappings between ontologies and inconsistency checking issues are shown.
- Existing integration obstacles due to the diversity of data formats and lack of formalization approaches as well as the trade-offs that are common in biological sciences.

- The knowledge will be weighted according to some defined evidence codes expressing the support media similar to those implemented in GO (experimental, electronically inferred, and so forth).
- Query system and graphical user interface are foreseen. The ultimate aim of the project is to support hypothesis evaluation about cell-cycle regulation issues.

Acknowledgements

This work was financially supported by the DIAMONDS project that is funded by the EU, contract number LSHG-CT-2004-512143.