Event Ontology: A Pathway-Centric Ontology for Biological Processes

Tatsuya Kushida, Toshihisa Takagi, and Ken Ichiro Fukuda

Pacific Symposium on Biocomputing 11:152-163(2006)

EVENT ONTOLOGY: A PATHWAY-CENTRIC ONTOLOGY FOR BIOLOGICAL PROCESSES

TATSUYA KUSHIDA*

Institute for Bioinformatics Research and Development,Japan Science and Technology Agency, AIST Bio-IT Research Building, 2-42 Aomi, Koutou-ku, Tokyo 135-0064, JAPAN

TOSHIHISA TAKAGI †

Graduate School of Frontier Sciences, University of Tokyo, 5-1-5 Kashiwanoha, Kashiwa-shi, Chiba, 277-8562, JAPAN

KEN ICHIRO FUKUDA‡

Computational Biology Research Center, National Institute of Advanced Industrial Science and Technology, AIST Bio-IT Research Building, 2-42 Aomi, Koutou-ku, Tokyo 135-0064, JAPAN

Event ontology is a new biomedical ontology developed to annotate pathway components in a pathway database. It organizes the concepts and terms of sub-pathways, pathways, biological phenomena, experimental conditions, medications, and external stimuli appearing in biological pathways (*e.g.* signal transduction, disease-, metabolic-, molecular interaction-, genetic interaction pathways, *etc.*). Concepts in the Event ontology are extracted manually from scientific literature. Each term has links to external databases such as Gene Ontology, Reactome, KEGG, BioCyc, and PubMed.

1. Introduction

Pathway databases are becoming increasingly important for the elucidation of mechanisms that underlie various types of biological phenomena, including disease. A number of pathway databases achieve these aims. For example, "KEGG"[1] and "BioCyc"[2] manage metabolic pathways in various organisms; "Reactome"[3] and "INOH"[4] attempt to represent various types of biological events such as immune response- and gene expression mechanisms as biological

* kushida-tatsuya@aist.go.jp
† tt@k.u-tokyo.ac.jp
‡ fukuda-cbrc@aist.go.jp

pathways; "aMaze",[5] "Patika",[6] and "Biocarta"[7] manage signal transduction- and disease pathways, and "BIND",[8] "DIP",[9] and "IntAct"[10] are protein–protein interaction databases. "Cytoscape"[11] and "GenMAPP"[12] are tools that utilize pathway data for data analysis. "BioPAX"[13] is a community project in which members engaged in the development of these pathway databases and tools cooperate to establish a data-exchange format for biological pathway data.

While conventional pathway data are highly structured, they use open vocabularies to name a pathway object. As this hampers data integration or mapping between data from different sources, the community needs a methodology to annotate data in a coherent way.

To solve this problem, we propose to use a set of annotation ontologies for pathway data annotation. In the INOH database, pathway diagrams are represented by compound graphs[14] and each pathway component is annotated by several biological ontologies. For example, protein objects are annotated by the MoleculeRole ontology[15] that manages the relations among generic protein names and concrete protein names that appear in the scientific literature. Consequently, INOH provides high-quality pathway data that allow advanced ontological searches, thereby improving accuracy by extending the search range using ontological trees.

Gene Ontology manages biological phenomena divided into "biological process", "molecular function", and "cellular component" categories and is generally used to annotate the functions of genes and gene products.[16] However, it is not designed to cope with the relations between pathways and sub-pathways and between pathways and their related biological phenomena. For example, Gene Ontology contains about 200 terms corresponding to "pathways", but few pathways include their sub-pathways with a part-of relationship [*e.g.*, transforming growth factor beta receptor signaling pathway (GO:0007179), I-kappaB kinase/NF-kappaB cascade (GO:0007249), and JAK-STAT cascade (GO:0007259)].

Furthermore, concepts regarding biological phenomena managed by Gene Ontology are too large and exhaustive for pathway data annotation. Terms, such as "actin filament-based process (GO:0030029)," seldom appear in articles that discuss biological pathways and are not used for pathway annotation.

GO slims are cut-down versions of Gene Ontology; they contain a subset of the terms included in Gene Ontology.[17] However, the terms in GO slims regarding pathway-related biological phenomena are insufficient.

In BioCyc, one of the most popular metabolic pathway databases, each "pathway" consists of "reactions" whose relations are classified hierarchically.[2] However, knowledge regarding signal transduction pathways is not rich enough to cover pathways in general.

Event ontology is designed to satisfy the need for a new pathway-centric biomedical ontology to annotate sub-pathways, pathways (biological processes), and their related biological phenomena in pathway databases. The knowledge was extracted from the scientific literature by manual expert curation. This ontology is expected to provide a structured high-quality controlled vocabulary for pathway data. The use of this ontology, instead of free text, for pathway or sub-pathway annotations facilitates advanced ontological searches and eases the pathway-annotation task of curators.

2. Curation Methods

Terms were collected by curators with a molecular biology background from the text and illustrations of review articles and original papers published in biomedical journals. In the present study, information on biological pathways in mammals (human and non-human) was curated. Event ontology manages the following concepts: (1) sub-pathways, (2) pathways, (3) biological phenomena related to these pathways, (4) experimental conditions under which pathways are investigated, and (5) external stimuli inducing a specific pathway. The organized terms are used to annotate sub-pathways, pathways, biological phenomena, experimental conditions, and external stimuli.

Event ontology was developed using DAG-Edit.[18] Terms are structured as a DAG (directed acyclic graph) and two relations, "is-a" and "part-of", are used. For each term, PubMed IDs are entered as reference information in the General Dbxrefs column in the text-editor panel. Sentences cited from the literature are entered in the Definition column. If there are corresponding terms in Gene Ontology[16], KEGG[1], BioCyc[2], Reactome[3], *etc.*, the IDs are entered in the General Dbxrefs. Synonyms for the terms are entered in the Synonyms column. In addition, if terms related to specific proteins or chemicals are found, their corresponding IDs in the MoleculeRole ontology[15] are entered in the General Dbxrefs column (Figure 1).

In the case of concepts investigated in only a specific organism, its name is recorded after each term name. Terms without a specific organism name indicate generic concepts that are investigated extensively in various organisms. Specific terms (*e.g.*, "Wnt signaling pathway [Mouse]") are entered under the corresponding canonical term (*e.g.*, "Wnt signaling pathway") with an "is-a" relationship.

Event ontology is distributed as an OBO file and can be converted to OWL. Therefore, it is useful for the application of/to semantic web technologies.

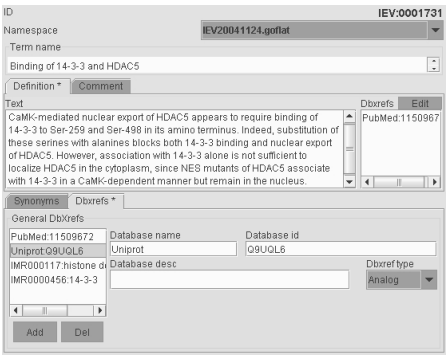

Figure 1. The text editor panel of DAG-Edit

3. Structure of Event ontology

3.1. *Biological event and Environmental event*

The ontology root has two subclasses, "Biological event (IEV:0000003)[§]", which includes sub-pathways (*e.g.*, "Binding of Smad3 and PIASy" and "MAPKKK cascade"), pathways (*e.g.,* Wnt signaling), and biological phenomena (*e.g.,* cell growth), and "Environmental event (IEV:0000185)", which includes experimental conditions (*e.g.,* medium), drugs/medications (*e.g.,* FK506 medication), and external stimuli (*e.g.,* ultraviolet irradiation).

3.2. *Pathways and Sub-pathways*

This is an example of how we manipulated the knowledge resources of Gene Ontology during the construction of Event ontology. We preserved the mapping between Event ontology and Gene Ontology if the latter contained a corresponding term, and we designed the DAG structure of terms by following the Gene Ontology conventions (*e.g.,*. "regulation", "negative regulation"). Furthermore, we collected just enough terms so as not to reduce pathway-curation efficiency.

As Gene Ontology does not include sufficient information regarding the relations between sub-pathways and pathways, we constructed these relations thoroughly in Event ontology. As in Gene Ontology, a sub-pathway is located under the pathway with a "part-of" relationship. For example, "Binding of

[§] "IEV" is the ID prefix of the Event ontology.

antigen and BCR complex (IEV:0001278)" is part-of "B cell receptor signaling (through IKK-NF-kappaB cascade) (IEV:0001298)" (Figure 2).

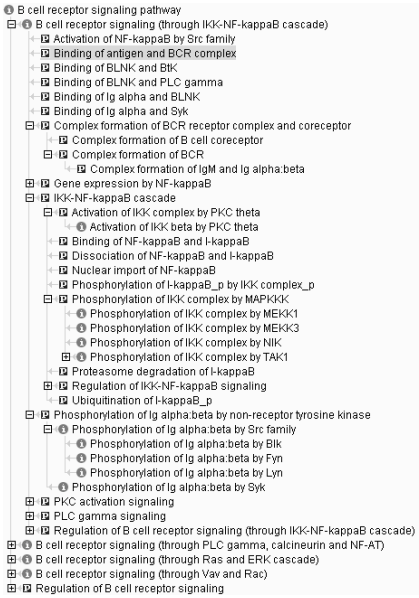

Figure 2. The relations between B-cell receptor signaling and the sub-pathways. The symbol "i" represents an "is-a" relationship; "P" indicates a "part-of" relationship.

3.3. *Types of sub-pathways*

In Event ontology, a sub-pathway is defined as a reaction in enzymatic reactions, an association or dissociation between molecules, or as a translocation of molecules. A sub-pathway in Event ontology corresponds with a "reaction" and a part-of "pathway" in the BioCyc[2] or Reactome.[3]

The sub-pathways are classified by molecular interaction types. Event ontology has about 30 molecular interaction types such as "Binding", "Phosphorylation", "Nuclear import", and "Deacetylation". "Deacetylation (IEV:0001681)" is-a "Hydrolysis (IEV:0000184)". This is-a relationship is used to define "kind-of" as in the further specialization of a class.

3.4. *Classification of sub-pathways using MoleculeRole ontology*

Each term name of a sub-pathway has a compositional structure and is based on molecular interaction types and the relevant molecule names. For molecular

names (protein and chemical names), terms are defined in MoleculeRole Ontology,[15] an ontology of protein names that appear in the scientific literature. It manages the relations among "*function names*" and "*abstract (generic) names*" and "*concrete names*". Thus, each term name of a sub-pathway in Event ontology may consist of a "*molecular function name*" such as "protein serine/threonine kinase", an "*abstract (generic) molecule name*" such as "MAPK", or a "*concrete molecule name*" such as "p38" (*e.g.*, "Phosphorylation of p38 by MKK3/6" (IEV:0000235)).

In Event ontology, "Phosphorylation of MKK4/7 by ASK1 (IEV:0000811)" is located under "Phosphorylation of MKK4/7 by MAPKKK (IEV:0000237)" with an "is-a" relationship. This is-a relationship is used to define specific instances of a class of protein. Thus, the vertical relations among the same molecular interaction type are determined by the relative degree of abstraction of the molecules constituting the sub-pathways.

3.5. *Cellular localization of Sub-pathways*

Furthermore, each sub-pathway term is classified according to its cellular localization. For example, "Phosphorylation in cytosol (IEV:0000025)" is located under "Phosphorylation (IEV:0000005)" with an "is-a" relationship. This is-a relationship is used to define different locations. On the other hand, "Phosphorylation of IKK complex by MAPKKK (IEV:0000250)" is located under "Phosphorylation in cytosol (IEV:0000025)" with an "is-a" relationship. This is-a relationship is used to define participants in processes. Thus, all sub-pathways are classified according to the relations between sub-pathways and pathways, the molecular interaction types, and cellular localization. Usage and the types of terms representing the cellular localization differ according to the molecular interaction types.

In the case of a direct interaction such as "Binding" and a metabolic reaction such as "Phosphorylation", one term, representing the cellular localization where the molecular interaction occurs, is used (*e.g.*, "Phosphorylation in nucleus (IEV: 0000231)"). On the other hand, in the case of a molecular translocation such as "Nuclear export", two terms indicating the start and end-point of the translocation are used [*e.g.*, "Translocation of Bax from mitochondrial membrane to cytosol (IEV:0001692)" and "Translocation of Calcium ion from ER to cytosol through calcium ion channel (IEV:0000274)"].

In Reactome,[3] "Cellular component" is used to represent the cellular localization of "Events" and "EntityCellular compartment" is used to represent the cellular localization of "molecules". The "EntityCellular compartment" is part of the "Cellular component". Furthermore, the "Cellular component" of

Reactome is part of "Cellular compartment" of Gene Ontology. However, we ascertained that the cellular localization information of Events was inconsistent in Reactome. For example, the cellular localization of the reaction "Internalization of the insulin receptor (ReacomeID:74718)" and "Translocation of tBID to mitochondria (RactomeID:139920)" is "cell" and "cytosol", respectively. Therefore, it is not reasonable to represent the cellular localization of the Event (*i.e.*, sub-pathway and pathway) by using a single term from "Cellular compartment" of Gene Ontology.

3.6. *Classification of biological phenomena related to pathways*

Biological phenomena related to biological pathways such as signal transduction were extracted from the literature and then classified into 4 levels of phenomena: "Molecular event (IEV:0000071)", "Cellular event (IEV:0000069)", "Organism event (IEV:0000082)", and "Physiological event (IEV:0001330)". "Cellular event" and "Physiological event" of Event ontology correspond to "cellular process (GO:0009987)" and "physiological process (GO:0007582)" of Gene Ontology, respectively.

3.7. *Relations between pathways and biological phenomena*

Event ontology manages the relations among a biological phenomenon (including disease) and the related pathway(s). In Event ontology, a pathway involved in a specific biological phenomenon is located under the corresponding biological phenomenon with a "part-of" relationship.

For example, the sentence, "The BCR induces the signals that are required for survival and proliferation of B cells. These include activation of PI3K-regulated AKT, RAS–RAF–ERK (extracellular signal-regulated kinase) and NF-B pathways"[19] is transcribed in Event ontology as: "B cell proliferation (IEV:0001562)" as a biological phenomenon is located above "B cell receptor signaling (through IKK-NF-kappaB cascade), (IEV:0001298)", "B cell receptor signaling (through Ras and ERK cascade), (IEV:0001295)", and "B cell receptor signaling (through Vav and Rac), (IEV:0001296)" as a pathway.

3.8. *Treatments & Media*

Event ontology manages the concepts of experimental conditions and abiotic external stimuli that induce a pathway and lead to a specific biological phenomenon as an "Environmental event (IEV:0000185)" with two subclasses, "Medium condition (IEV:0000293)" and "Treatment (IEV:0000387)". For example, the "Medium condition" class includes "BME medium

(IEV:0000313)", the "Treatment" class includes "Anticancer drug medication (IEV:0001566)", "Ultraviolet irradiation (IEV:0001568)", "Hydrogen Peroxide treatment (IEV:0001573)", and "Heat shock treatment (IEV:0001565)". The sentence: "Caloric restriction activates cell survival signaling and then induces cell survival"[20] is expressed in Event ontology as "Caloric restriction (IEV:0001678)" is a "Treatment" and a part of "Caloric restriction cell survival signaling (IEV:0001686)" that is a part of "Cell survival (Inhibition of apoptosis) (IEV:0000153)".

4. Application of Event ontology

4.1. *Protein and Gene annotation using Event ontology*

Event ontology manages the relations among sub-pathways, pathways, and biological phenomena. The terms representing sub-pathways contain molecule names controlled by MoleculeRole ontology (*e.g*, "R-smad") and molecular interaction types (*e.g*, "Binding"). Furthermore, the terms have links to external references, such as Gene Ontology,[16] Reactome,[3] KEGG,[1] BioCyc,[2] MoleculeRole ontology,[15] and PubMed.

Thus, to a researcher interested in a gene or protein, querying the molecule name using Event ontology returns information regarding all sub-pathways involving the molecule and provides links to the actual pathway data annotated by Event ontology terms. Furthermore, knowledge regarding related biological phenomena such as apoptosis and immune responses can be acquired with Event ontology from the relations among sub-pathways and pathways, and from relations among pathways and biological phenomena. By annotating proteins or genes using Event ontology, the molecules can be linked to pathway information and pathway data.

In this sense, as Event ontology connects molecules to various levels of biological phenomena via sub-pathways and pathways in the ontological structure, it facilitates the seamless navigation among the concepts of molecules, sub-pathways, pathways, and biological phenomena.

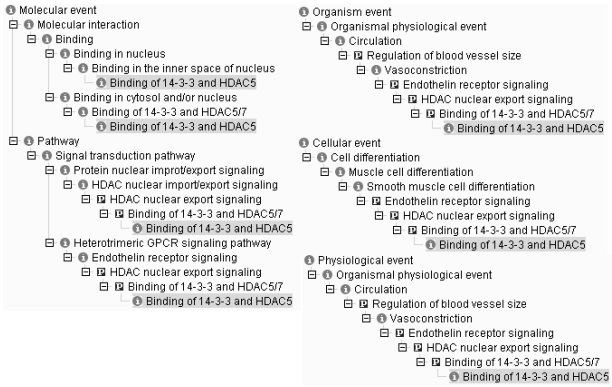

Figure 3. The pathways and biological phenomena involved in the sub-pathway "Binding of 14-3-3 and HDAC5"

4.2. *Example of an ontological search using Event ontology*

To demonstrate a query with Event ontology, we present a search example using DAG-Edit.[18] When we entered the query term "14-3-3" in the query column of DAG-Edit, 14 sub-pathways were found (*e.g.*, "Binding of 14-3-3 and HDAC5, (IEV:0001731)" and "Gene expression of 14-3-3 by p53 (IEV:0001112)"). We selected "Binding of 14-3-3 and HDAC5". The DAG viewer of DAG-Edit showed the cellular localization of "Binding of 14-3-3 and HDAC5" to be the "cytosol" or "within the nucleus" (Figure 3) and indicated that it was a sub-pathway of "Endothelin receptor signaling (IEV:0001693)". The pathway was shown to be involved in "vasoconstriction (IEV:0001822, GO:0042310)" and "muscle cell differentiation (IEV:0001824, GO:0051145)". In addition, PubMedIDs were provided as reference information in Dbxrefs of the DAG-Edit. "Binding of 14-3-3 and HDAC5" was also identified as a sub-pathway of "HDAC nuclear export signaling (IEV:0001866)".

The definition and Dbxrefs of the term "Binding of 14-3-3 and HDAC5, (IEV:0001731)", *e.g.*, PubMedID:11509672 and Uniprot:Q9UQL6 (Figure 1) provide comprehensive information to the user that HDAC5 is a deacetylation enzyme, and the nuclear export of HDAC5 as a result of 14-3-3 and HDAC5 binding leads to the inhibition of deacetylation of histones in nuclear DNA, and that the event controls expression of some gene(s).

5. Conclusions

5.1. *Comparison between Event ontology and other ontologies*

Before developing Event ontology, we investigated whether available ontologies were reasonable to annotate the pathways and sub-pathways in INOH[4] pathway diagrams. We concluded that Gene Ontology,[16] the most widely used biomedical ontology for gene function annotation, was not suitable for pathway data annotation because (1) it does not manage thoroughly the relations among sub-pathways and pathways, (2) it does not manage the relations between pathways and related biological phenomena, and (3) its set terms are too large and exhaustive for annotation of the pathway components.

GO slims (*e.g.*, Generic GO slim) are particularly useful for providing a summary of the results of GO annotations of a genome, microarray data, or cDNA collection when a broad classification of gene product function is required. However, the terms regarding the biological phenomena related to pathways are insufficient in GO slims. For example, Generic GO slim does not have any terms for pathway-related biological phenomena such as "actin filament organization (IEV:0001324, GO:0007015)" or "stress fiber formation (IEV:0000095, GO:0043149)" that are led, for example, by "Integrin signaling pathway (through Rho) (IEV:0000616)" and "TGF beta super family signaling pathway (IEV:0000090)".

Event ontology is a pathway-centric complement to the GO biological process ontology. We carefully followed the Gene Ontology conventions to structure our terms. Since there is a certain amount of complementarity, we are planning to submit our terms to the Gene Ontology consortium.

In Reactome,[3] each "reaction" or "event" is not annotated with controlled vocabularies. For example, terms such as "translocation", "transport", and "internalization" are used to represent the translocation of molecules. The definitions of these vocabularies and relations are not recorded in Reactome [*e.g.*, "Translocation of BIM to mitochondria [Homo sapiens] (ReactomeID:139919)", "Notch 2 precursor transport to golgi [Homo sapiens] (ReactomeID:157077)", and "Internalisation of the insulin receptor [Homo sapiens] (ReactomeID:74718)"]. This complicates retrieval by the system of all translocation-related reactions.

In BioCyc,[2] each "pathway" consists of "reactions" and the relations are classified hierarchically with the hierarchy mainly covering metabolic pathways. We are planning to add the entire metabolic pathway classification of BioCyc under the "Metabolic pathway (IEV:0000818)" term in Event ontology.

Event ontology attempts to manage concepts of various types of biological pathways such as signal transduction, metabolic-, molecular interaction-, genetic interaction-, and disease pathways, and it yields thorough information concerning the relations among sub-pathways and pathways and among pathways and related biological phenomena. Event ontology is a new biomedical ontology to annotate various types of biological pathway components.

5.2. *Statistics*

Table 1 shows the latest statistics of Event ontology. Currently, the number of total terms, at 2289, is about 1/8 of Gene Ontology (19,081 in June 2005).[16] We selected the terms in Event ontology carefully to achieve effective pathway annotation. Event ontology is an ongoing project and is updated on a regular basis.

Event ontology and MoleculeRole ontology in OBO format can be downloaded from our project web site (http://www.inoh.org/).

Table 1. The latest statistics of Event ontology (June 30, 2005). *: "Others" includes class names to classify concepts.

Pathways	*e.g.*, Wnt signaling	121
Sub-pathways	*e.g.*, Binding of Wnt and Frizzled	1794
Biological phenomena	*e.g.*, Apoptosis	168
Environmental events	*e.g.*, FK506 medication	127
Others*	*e.g.*, Binding in cytosol	79
Total		2289

5.3. *Future work*

Event ontology has a compositional structure. Mapping terms from other ontologies such as MoleculeRole ontology and Cell location ontology will enrich the ontology and the computation of implicit relations from these term relations represents an interesting research topic. We are part of the BioPAX working group and strongly believe Event ontology should incorporate existing communitiy-based standards, *e.g.,* BioPAX and OBO. The current Event ontology is limited to provide a structured controlled vocabulary for (pathway) data annotations. Decomposing the current "is-a" relation into more explicit relations such as "kind_of", "located_in" and "has_participant" relations[21] is the subject of ongoing work in our laboratory.

Acknowledgments

This work was supported by BIRD of Japan Science and Technology Agency.

References

1. M. Kanehisa and S. Goto, *Nucleic Acids Res*. **28**, 27-30 (2000).
2. CJ. Krieger, P. Zhang1, LA. Mueller, A. Wang, S.e Paley, M. Arnaud, J. Pick, S. Y. R. and PD. Karp, *Nucleic Acids Res*. **32**, 438-442 (2004).
3. G. Joshi-Tope, M. Gillespie, I. Vastrik, P. D'Eustachio, E. Schmidt, B. de Bono, B. Jassal, GR. Gopinath, GR. Wu, L. Matthews, S. Lewis, E. Birney and L. Stein, *Nucleic Acids Res*. **33**, D428-432 (2005).
4. INOH database website: http://www.inoh.org/
5. C. Lemer, E. Antezana, F. Couche, F. Fays, X. Santolaria, R. Janky, Y. Deville1, J. Richelle and SJ. Wodak, *Nucleic Acids Res*, **32**, 443–448 (2004).
6. E. Demir, O. Babur, U. Dogrusoz, A. Gursoy, A. Ayaz, G. Gulesir, G. Nisanci and R and Cetin-Atalay, *Bioinformatics* **20**, 349–356 (2004).
7. BioCarta website: http://www.biocarta.com/
8. BIND website: http://bind.ca/
9. DIP website: http://dip.doe-mbi.ucla.edu/
10. IntAct website: http://www.ebi.ac.uk/intact/
11. P. Shannon, A. Markiel1, O. Ozier, NS. Baliga, JT. Wang, D. Ramage, N. Amin, B. Schwikowski and T. Ideker, *Genome Research* **13**, 2498-2504 (2003).
12. KD. Dahlquist, N. Salomonis, K. Vranizan, SC. Lawlor, BR. Conklin, *Nat Genet*. **31**, 19-20 (2002).
13. BioPAX website: http://www.biopax.org/
14. K. Fukuda and T. Takagi, *Bioinformatics* **17**, 829-837 (2001).
15. S. Yamamoto, T. Asanuma, T. Takagi and K. Fukuda, *Comparative and Functional Genomics* **5**, 528-536 (2004).
16. The Gene Ontology Consortium, *Genome Res*. **11**, 1425-1433 (2001).
17. GO Slims website: http://www.geneontology.org/GO.slims.shtml
18. DAG-Edit website: http://sourceforge.net/projects/geneontology
19. H. Niiro H and EA. Clark, *Nature Reviews Immunology* **2**, 945-956 (2002).
20. HY. Cohen, C. Miller, KJ. Bitterman, NR. Wall, B. Hekking, B. Kessler, KT.Howitz, M. Gorospe, R. de Cabo, DA. Sinclair, *Science* **305**, 390-392 (2004)
21. B. Smith, W. Ceusters, B. Klagges, J. Kohler, A. Kumar, J. Lomax, C. Mungall, F. Neuhaus, AL. Rector and C. Rosse, *Genome Biology* **6**, R46 (2005)