In Silico Biology 5 (2005) 1–3
IOS Press

Preface

Third "Ontology Workshop on Ontology and Genomes"

Takako Takai-Igarashi[a], Toshihisa Takagi[b], Holger Michael[c] and Edgar Wingender[c,d]
[a]*Graduate School of Information Science and Technology, University of Tokyo, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-0033, Japan*
E-mail: taka@bi.is.s.u-tokyo.ac.jp
[b]*Department of Computational Biology, Graduate School of Frontier Sciences, University of Tokyo, 5-1-5, Kashiwanoha, Kashiwa-shi, Chiba, 277-8562, Japan*
E-mail: tt@k.u-tokyo.ac.jp
[c]*Department of Bioinformatics, Medical School, University of Göttingen, Goldschmidtstr. 1, D-37077 Göttingen, Germany*
E-mail: {hom, e.wingender}@bioinf.med.uni-goettingen.de,
[d]*BIOBASE GmbH, Halchtersche Str. 33, D-38304 Wolfenbüttel, Germany*
E-mail: ewi@biobase.de

INTRODUCTION

The third "Ontology Workshop on Ontology and Genomes" (http://www.bioinf.med.uni-goettingen.de/ontology-workshop/), held at Goettingen University in July 1–3, 2004, was devoted to the development of ontologies for OMICS research and their application in either of these fields. It was the idea of the workshop to bring together experts in genomics, transcriptomics, proteomics or metabolomics with specialists in conceptualizing, developing and maintaining ontologies in order to find out more about the ontological needs OMICS researchers may have, and which of these requirements might be fulfilled by ontologists now or in future. Accordingly, the program comprised two sessions about Ontologies which framed the sessions on Genomics, Proteomics/Protein Names & Functions, Transcriptomics/Cytomics, Metabolomics, and Proteomics/Protein Networks.

The introductory session on Ontologies was opened by a keynote lecture of J. Bard (Edinburgh, UK), followed by S. Schulze-Kremer (Univ. Hannover, Germany) and I. Rojas (Heidelberg, Germany), who introduced a number of basic principles of ontologies and how they can be applied to life sciences in general or, more specifically, to structures (anatomy) and functions of biological systems.

The field of Genomics may be the most established Omics area, thus the contribution of the keynote given by M.-P. Lefranc (Montpellier, France) focused on one particular, though probably one of the most important functional aspects: the linkage between immunogenetics and immunoinformatics [1].

Electronic publication can be found in *In Silico Biol.* **5**, 0001 <http://www.bioinfo.de/isb/2004/05/0001/>.

Concepts about how to store and handle data on genome-encoded biological networks using ontologies, thereby linking already the different Omics areas, were introduced in this session by J. Koehler (Harpenden, UK) [2].

From the session Proteomics/Protein Names & Functions, two out of three talks given by K. Satou and A. Koike are reflected in this Special. The corresponding articles specify how the literature relevant for Proteomics studies can be automatically extracted using controlled vocabularies or ontologies [3,4], making use but also exceeding the scope of Gene Ontology (GO). In doing so, it becomes increasingly obvious that the required knowledge for understanding biology refers not only to objects such as genes and proteins, but to pathways and networks such as protein interaction networks [4]. In addition, O. Hoffmann demonstrated his approach to automatically extract disease-relevant information about enzymes from PubMed abstracts which have been incorporated in the BRENDA database, later on reviewed in the metabolomics session (see below).

"Transcriptomics" became nearly a synonym for systematic gene expression analysis, and therefore has been combined in one session with "Cytomics", the systematic and comprehensive analysis of the cellular network of an organism (its "Cytome"), preferably the human and model organisms. In his keynote lecture to this session, A. Kel raised the general subject about the nature of explanatory models in molecular biology and to what extend ontologies may help in finding explanations to observed phenomena (see below for further discussion). J. Landgrebe gave an excursus about the statistical analysis of microarray expression data, which resulted in an article beyond this Special [5], and K. Okubo presented how to make use of textbook structured vocabularies for biomedical knowledge extraction. A way to derive an ontology for Cytomics was discussed by H. Michael and is described in a Communication [6].

The already well-established research area of Metabolomics was introduced by a keynote from D. Schomburg, Univ. of Köln, explaining the complex interlinks between experimental setups and needs on the one and the corresponding database and algorithmic requirements on the other side. A more recent variant was introduced by S. Goto (Kyoto) for the field of Glycomics.

The second session purely devoted to Ontologies dealt with the great contributions of Gene Ontology (GO) to the diverse fields of Omics research through the Open Biology Ontologies (OBO) resources, as presented by M. Harris, as well as with the GOA interface between GO and the UniProt Knowledgebase described by V. Lee [7]. R. Stevens discussed the use of biological and bioinformatics ontologies in integrating data and results from in silico experiments, as done in the myGrid project. Finally in this session, the use of ontologies for NLP (Natural Language Processing) approaches was presented by J. Saric. This talk also became part of the selection for this Special [8].

The final session about Proteomics and, or rather, on protein and signaling networks, set a promising punch line behind the presentations of the Workshop. S. Kuroda described systems analysis of signaling networks and novel regulatory mechanisms in the ERK cascade, T. Takai-Igarashi presented her Cell Signaling Network Ontology (CSNO) for representing biological functions of signaling pathways, and E. Wingender described the hierarchical organization of signaling components and reactions in the TRANSPATH database.

In two discussion rounds, one evening Round Table Discussion and the final discussion after the last talks, the participants tried hard to make clear among themselves what Ontologies are in contrast to taxonomies, general classifications etc., though agreed also that it is of limited value to spend endless efforts in this kind of definitory debates. The goal was to come to a mutual understanding about the basic concepts in genomics, proteomics (sensu interaction networks), transcriptomics, metabolomics, etc. and their global approaches to characterize biological systems. Questions raised during the presentations and the discussions were: Is it achievable and still consistent with the properties of an ontology to quantify

the relationships of its entities, e.g. by adding similarity measures? And is it conceivable to model causal relationships in the ontologies we are using already or that will be in use in near future? What seems to be needed from the side of Omics applications are links from the observed facts to the explanatory models that lead to decision trees. What we would need, then, are ontologies of the components of the explanatory models. This would be consistent with S. Schulze-Kremer's statement that a knowledge base is an ontology plus facts (or "factoids"), and that ontologies should enable us to reconstruct physical phenomena.

ACKNOWLEDGEMENTS

The organizers are indebted to the Japanese Ministry of Education, Culture, Sports, Science and Technology (MEXT), Volkswagen Foundation, BIOBASE GmbH and to the University of Göttingen for financial support.

REFERENCES

[1] Lefranc, M.-P., Clément, O., Kaas, Q., Duprat, E., Chastellan, P., Coelho, I., Combres, K., Ginestoux, C., Giudicelli, V., Chaume, D. and Lefranc, G. (2004). IMGT-Choreography for immunogenetics and immunoinformatics. In Silico Biol. **5**, 0006.

[2] Koehler, J., Rawlings, C., Verrier, P., Mitchell, R., Skusa, A., Ruegg, A. and Philippi, S. (2004). Linking experimental results, biological networks and sequence analysis methods using Ontologies and Generalised Data Structures. In Silico Biol. **5**, 0005.

[3] Satou, K. and Yamamoto, K. (2004). Utilizing weakly controlled vocabulary for sentence segmentation in biomedical literature. In Silico Biol. **5**, 0008.

[4] Koike, A. and Takagi, T. (2004). PRIME: automatically extracted PRotein Interactions and Molecular Information databasE. In Silico Biol. **5**, 0003.

[5] Landgrebe, J., Bretz, F. and Brunner, E. (2004). Efficient two-sample designs for microarray experiments with biological replications. In Silico Biol. **4**, 0038.

[6] Michael, H. Chen, X., Fricke, E., Haubrock, M., Ricanek, R. and Wingender, E. (2004). Deriving an ontology for human gene expression sources from the CYTOMER® database on human organs and cell types. In Silico Biol. **5**, 0007.

[7] Lee, V., Camon, E., Dimmer, E., Barrell, D. and Apweiler, R. (2004). Who tangos with GOA? - Use of Gene Ontology Annotation (GOA) for biological interpretation of '-omics' data and for validation of automatic annotation tools. In Silico Biol. **5**, 0002.

[8] Saric, J., Jensen, L. J. and Rojas, I. (2004). Large-scale extraction of gene regulation for model organisms in an ontological context. In Silico Biol. **5**, 0004.