Christian J. Stoeckert
Research Associate Professor
1415 Blockley Hall
423 Guardian Drive
Philadelphia, PA 19104-6021
Tel: 215.573.4409
Fax: 215.573.3111
e-mail: stoeckrt@pcbi.upenn.edu

March 21, 2003

Daniel Drel, Ph.D.
Department of Energy
Life Sciences Divisions, SC-72
Office of Biological Research, Office of Environmental Science
Germantown, MD 20874-1290

Department of Energy
Chicago Operations Office
9800 South Cass Avenue
Argonne, Illinois 60439

RE: DE-FGO2-00ER62893
Dear Dan:

DOE Patent Clearance Granted
MP Dvorscak
Mark P. Dvorscak
(630) 252-2393
E-mail: mark.dvorscak@ch.doe.gov
Office of Intellectual Property Law
DOE Chicago Operations Office
Date Apr 11 2005

Please find enclosed publications resulting from the "Information Management Infrastructure for the Systematic Annotation of Vertebrate Genomes" project along with a brief summary.

The primary accomplishment of the project has been the development of a functional genomics database system, GUS (Genomics Unified Schema). The system provides a relational schema, application framework, and web interface for multiple projects and includes information on sequences, gene expression, gene regulation, ontologies, and data provenance. The system has been made portable and has been set up at other institutions. A web site is available describing the system and providing links for downloading code at <http://www.gusdb.org>. A description of the initial version of GUS was published in:

Davidson, S.B., Crabtree, J., Brunk, B., Schug, J., Tannen, V., Overton, G.C., Stoeckert, C.J.Jr. (2001) K2/Kleisli and GUS: Experiments in integrated access to genomic data sources. IBM Systems Journal. 40:512-531.

The GUS system has been applied toward the annotation of the human and mouse genome using the DoTS (Database of Transcribed Sequences) component. DoTS are assemblies of transcribed sequences (ESTs and mRNAs) that can be used to generate gene models on genomic sequence and integrate annotations. A description of the use of DoTS to both annotate a region of mouse chromosome 5 and also assemble genomic contigs was published in:

Crabtree, J., Wiltshire, T., Brunk, B., Zhao, S., Schug, J., Stoeckert, C.J.Jr., Bucan, M. (2001) High-resolution BAC-based map of the central portion of mouse chromosome 5. Genome Res. 11:1746-57.

The GUS system has also provided the opportunity to use the integrated information stored in GUS to develop algorithms for data mining. For many genomes including human

DISCLAIMER

This report was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor any agency Thereof, nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or any agency thereof. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or any agency thereof.

DISCLAIMER

Portions of this document may be illegible in electronic image products. Images are produced from the best available original document.

and mouse the function of genes is unknown. A description for predicting gene function from analysis of Gene Ontology assignments and associated protein domains is published in:

Schug, J., Diskin, S., Mazzarelli, J, Brunk, B.P., Stoeckert, C.J.Jr.
(2002) Predicting Gene Ontology functions from ProDom and CDD protein domains.
Genome Research. 12:648-655.

Database interoperability is important for sharing information. An effort to create standards for microarray data resulted in a Microarray Gene Expression Object Model (MAGE-OM) and its XML implementation (MAGE-ML). A description of MAGE is published in:

Spellman, P.T., Miller, M., Stewart, J., Troup C., Sarkans, U.,
Chervitz, S., Bernhart, D., Sherlock, G., Ball, C., Lepage, M., Swiatek, M., Marks, W.L.,
Goncalves, J., Markel, S., Iordan, D., Shojatalab, M., Pizarro, A., White, J., Hubley, R.,
Deutsch, E., Senger, M., Aronow, Robinson, A., Bassett, D., Stoeckert, C.J. Jr., Brazma,
A. (2002) Design and implementation of microarray gene expression markup language
(MAGE-ML). Genome Biology 3(9):research0046.1-9.

Sincerely yours,

Christian J. Stoeckert, Jr., Ph.D.
Research Associate Professor