

- (2) a translator of OPM schemas into relational database specifications and procedures;
- (3) utilities for publishing OPM schemas in text (Latex), diagram (Postscript), and Html formats;
- (4) a translator of OPM queries into SQL queries;
- (5) a retrofitting tool for constructing OPM schemas (views) for existing relational genomic databases;
- (6) a tool for constructing Web-based form interfaces to MBDs that have an OPM schema; this tool was developed by Stan Letovsky at Johns Hopkins School of Medicine, as part of a collaboration.

The OPM data management tools have been highly successful in developing new genomic databases, such as GDB 6 (released in January 1996; <http://gdbgeneral.gdb.org/gdb/>) and the relational version of PDB (<http://terminator.pdb.bnl.gov:4148>), and in constructing OPM views and interfaces for existing genomic databases such as GSDB 2.0. The OPM data management tools are currently used by over ten groups in USA and Europe. The research underlying these tools is described in several papers published in scientific journals and presented at database and genome conferences.

In the past year the OPM tools have been presented at database and bioinformatics conferences, including the International Symposium on Theoretical and Computational Genome Research, Heidelberg, Germany, March 1996, the Workshop on Structuring Biological Information, Heidelberg, Germany, March 1996, the Meeting on Genome Mapping and Sequencing, Cold Spring Harbor, May 1996, the International Sybase User Group Conference, May 1996, the Bioinformatics -Structure Conference, Jerusalem, November 1996, and the Pacific Symposium on Bioinformatics, January 1997.

The results of the research and development underlying the OPM tools work have been presented in papers published in proceedings of database and bioinformatics conferences; these papers are available at <http://gizmo.lbl.gov/opm.html#Publications>.

DOE Contract No. DE-AC03-76SF00098.

## **The Genome Topographer: System Design**

S. Cozza, D. Cuddihy, R. Iwasaki, M. Mallison, C. Reed, J. Salit, A. Tracy, and T. Marr  
 Cold Spring Harbor Laboratory; Cold Spring Harbor, NY 11724  
 Marr: 516/367-8393, Fax: -8461, [marr@cshl.org](mailto:marr@cshl.org) or [marr@cb.cshl.org](mailto:marr@cb.cshl.org)

Genome Topographer (GT) is an advanced genome informatics system that has received joint funding from DOE and NIH over a number of years. DOE funding has focused on GT tools supporting computational genome analysis, principally on sequence analysis. GT is scheduled for public release next spring under the auspices of the Cold Spring Harbor Human Genome Informatics Research Resource. GT has 17 major existing frameworks: 1. Views, including printing, 2. Default manager, 3. Graphical User Interface, 4. Query, 5. Project Manager, 6. Workspace Manager, 7. Asynchronous Process Manager, 8. Study Manager, 9. Help, 10. Application, 11. Notification, 12. Security, 13. World Wide Web Interface, 14. NCBI, 15. Reader, 16. Writer, 17. External Database Interface. GT Frameworks are independent sets of VisualWorks (client) or SmallTalkDB (GemStone) classes which interact to perform the duties required to satisfy the responsibilities of the specific framework. Each framework is clearly defined and has a well-defined interface to use it. These frameworks are used over and over in GT to perform similar duties in different places. GT has basic tools and special tools. Basic tools get used many times in different applications, while special tools tend to be special purpose, designed to do fairly limited things, although the distinction is somewhat arbitrary. Tools typically use several frameworks when they get assembled. Basic Tools: 1. Project Browser, 2. Editor/Viewer, 3. Query, 4. NCBI Entrez, 5. File reader/writer, 6. Map comparison, 7. Database Administrator, 8. Login, 9. Default, 10. Help. Special Tools: 1. Study Manager, 2. Compute Server, 3. Sequence Analysis, 4. Genetic Analysis. These frameworks and tools are combined with a comprehensive database schema of very rich biological expression linked with pluggable computational tools. Taken together, these features allow users to construct, with relative ease, on-line databases of the primary data needed to study a genetic disease (or genes and phenotypes in general) from the stage of family collection and diagnostic ascertainment through cloning and functional analysis of candidate genes, including mutational analysis, expression information, and screening for biochemical interactions with candidate molecules. GT was designed on the premise that a highly informative, visual presentation of comprehensive data to a knowledgeable user is essential to their understanding. The advanced software engineering techniques that are promoted by using relatively new object oriented products has allowed GT to become a highly interactive and visually-oriented system that allows the user to concentrate on the problem rather than on the computer. Using the rich data representational features characteristic of this technology, the GT software enables users to construct models of real-world, complex biological phenomena. These unique features of GT are key to the thesis that such a system will allow users to discover otherwise intractable networks of interactions exhibited by complex genetic diseases.