



Sophic Systems Alliance, Inc.
Integrated Life Science Solutions

Press Release

FOR IMMEDIATE RELEASE- FOR IMMEDIATE RELEASE- FOR IMMEDIATE RELEASE

Sophic and Biomax Win Second NCI Cancer Gene Curation Project

Falmouth, MA – November 30, 2005. Sophic Systems Alliance Inc. today announced, together with Biomax Informatics AG, that Sophic signed a second project with the National Cancer Institute to text mine Medline abstracts and manually annotate and validate the results to identify all cancer genes, gene-disease relationships and cancer gene-compound relationships. The goal is to provide investigators with high quality information on cancer genes to accelerate their research.

“Last year we successfully completed the pilot phase of the cancer gene curation project,” said Patrick Blake, CEO of Sophic. “Our results demonstrated the unique power of BioLT™, the Biomax Linguistic System, to efficiently and accurately extract specific information from huge amounts of free text combined with the extraordinary capabilities of the Biomax annotation team to further refine the content.”

Dr. Kaj Alderman, the Biomax Project Leader said, “To prepare for the NCI pilot project, our linguistic team used BioLT to text mine over 15 million Medline abstracts to extract and build proprietary dictionaries. Publicly available cancer terms in the NCI Thesaurus and other ontologies were integrated with the BioLT dictionaries to create an extensive and accurate text mining foundation. The scope of the annotation project meant that we had to review and validate huge amounts of text in an assembly line process with the emphasis on quality control. We enhanced the Biomax BioXM™ Knowledge Management System annotation module to support the over 20 scientists on the annotation team.”

NCI specified Medline abstracts as the corpus for text mining which included over 8 million papers published since 1972 when the first oncogene was identified. In the automated step, BioLT was used by the scientists to search all of the abstracts, taking great care to make sure that no potential cancer genes were overlooked. Over 8,000 suspect cancer genes were identified and after manual review and validation, 4,500 “true” cancer genes met the strict criteria set by NCI for identifying a cancer gene. From the 4,500 “true” cancer genes, NCI identified 1,000 genes to continue through the entire curation process for the pilot.



Sophic Systems Alliance, Inc.
Integrated Life Science Solutions

“This was a test to see if the results of such a large scale text mining and manual annotation process could produce high quality results,” said Dr. Klaus Heumann, President of Biomax Informatics AG. “We used controlled vocabularies, NCI specified evidence and role codes to annotate and validate each of the 1,000 genes. This disciplined approach produced a consistent, high quality data set that will allow scientists to use automated search tools to mine the data for information about cancer.” He concluded with, “We are honored to support this vital effort.”

NCI confirmed their satisfaction with the results of the pilot by awarding the continuation of the cancer gene curation project to the Sophic-Biomax team. As the project moves forward, NCI plans to make the cancer genes publicly available every six months. Sophic will announce the availability of the genes in future press releases and provide a link from their website to the NCI URL where the genes can be accessed.

Heumann went on to say, “We integrated raw data in relational and flat-files a long time ago with our BioRS Data Integration and Retrieval System. Now, the challenge is to build on that foundation by integrating knowledge so scientists can create and see models of relationships, networks or pathways that they otherwise would have been hard pressed to see.” He concluded with, “BioXM is an agnostic, flexible and highly configurable knowledge framework designed to integrate semantic objects based on their scientific relationships. The next release of BioXM is scheduled for delivery sometime in Q4 2005.”

Blake concluded with, “At the September 30th caBIG Conference, Dr. Anna Baker, Deputy Director of NCI said that 1,500 Americans die each day of cancer. Dr. Ken Buetow, Director NCI Center for Bioinformatics reached out to commercial companies to become full partners in the caBIG community. Collaboration between government, academic and commercial partners to fight cancer is the only way forward.”

Biomax Informatics AG is implementing the NCI caBIG compatibility standards in their software development labs. After compliance has been certified by NCI, announcements will be made on the availability of caBIG compliant Biomax software. Sophic recently received its GSA Award allowing the company to be the prime integrator for this multi-year project. Additional details on the Cancer Gene Curation Project are provided in a white paper and a joint NCI-Biomax Poster on the Sophic Systems Alliance Inc. website. Go to www.sophicalliance.com

About Sophic Systems Alliance.



Sophic Systems Alliance, Inc.

Integrated Life Science Solutions

Sophic Systems Alliance Inc. (SSA) is an integrator of bioinformatics software development companies and provides a full range of professional services. SSA sells, markets, implements and supports the delivery of a broad range of software modules developed by Biomax Informatics AG, Geospiza, Inc. and other members of the alliance. The company is focused on providing professional services to scientists and agencies in Washington, D.C. including scientific consulting and custom bioinformatics software development. Customers include government research organizations, pharmaceuticals, and biotech companies. Sophic is an Optimized IBM Business Partner. More information is available at www.sophicalliance.com.

About Biomax Informatics AG.

Biomax, a recognized pioneer in the development of customized bioinformatics solutions, was founded in 1997 as a spin-off of the GSF-MIPS academic research group, now the German Research Center for Environment and Health-Institute for Bioinformatics (GSF-IBI). Founded by Prof. Dr. H. W. Mewes, Prof. Dr. D. Frishman and Dr. K. Heumann, Biomax developed a broad portfolio of bioinformatics solutions, including the well-known Pedant-Pro™ Sequence Analysis Suite, the BioRS™ Data Integration and Retrieval System, and other bioinformatics tools and services used in metabolic analysis, proteomics, and gene expression analysis. Additional information about Biomax can be found at the company's site on the World Wide Web at www.biomax.com.

Press Contact:

Chris Dozier
Sophic Systems Alliance, Inc.
(508) 495-3803
chris@sophicalliance.com



Sophic Systems Alliance, Inc.
Integrated Life Science Solutions

FACT SHEET

Sophic and Biomax Win Second NCI Cancer Gene Curation Project

1. NCI is encouraging commercial partners to leverage open source content and tools developed by caBIG team members.
2. Sophic and Biomax strategy is aligned with NCI caBIG to develop and deliver both content and applications to enhance and leverage the cancer gene database.
3. The first 1,000 cancer genes from the pilot have been combined with the manually annotated Biomax Human Genome Database.
4. As additional cancer gene information is produced, the new content will be integrated with the Human Genome Database.
5. Biomax AG has a proven track record of producing high quality, manually annotated genomic databases for over 8 years.
6. Use case demonstrations of the integrated Cancer Human Genome Database with the BioXM Knowledge Management System are focused on various types of cancer diseases, pathology information integrated with patient and clinical information.
7. The completion of the entire curation of the 4,500 cancer genes is planned for completion in Q4 of 2006.