



## Sophic to Use \$750K NCI Grant to Complete Cancer Biomarker Knowledgebase

October 15, 2010

By Uduak Grace Thomas

**Sophic Systems Alliance** has snagged a two-year, Phase II \$750,000 Small Business Innovation Research Grant from the National Cancer Institute that will fund the continued development of SCan-MarK — a knowledgebase that the company is intending as a one-stop shop for cancer biomarker information for oncologists, drug developers, and cancer researchers.

A beta version of the database is expected to be available next fall. When it is completed, SCan-MarK will contain curated and annotated documented biomarkers that have already been approved by the Food and Drug Administration, as well as candidate and target biomarkers that have been published in the scientific literature but are yet to be approved, Pat Blake, Sophic's CEO, told *BioInform*.

"A pipeline of all of the biomarkers is really what SCan-MarK is going to be," he said. "It's a moving, changing dynamic source of information that will be current, accurate, and up to date."

Under a Phase I SBIR awarded in 2008, which was worth \$150,000 and lasted six months, four Sophic scientists, working with collaborators at NCI, developed a prototype of the database, which Blake described as "almost a feasibility study in pulling together all the components" of the resource.

According to the [grant abstract](#) for the first phase of the study, Sophic planned to "review a sample of biomarker databases to determine the quality of the data, identify common data elements, and design a 'biomarker object' model that can be integrated into a biomarker knowledgebase."

In addition, the team planned to "add end-user query, mining, modeling and visualization tools" to help cancer researchers and clinicians "find and use up-to-date biomarker information."

For the second phase of the project, which is expected to last two years, Sophic's scientists plan to build on the prototype developed in phase one by collecting biomarker

data for five cancer types: breast, ovarian, colon, melanoma, and non-Hodgkin lymphoma.

The Phase II funding allowed Sophic to hire three new scientists, bringing the total number of its employees to 10 including full-time staff and consultants.

"We will go through the process of validating the individual genes and gene panels that are associated with those diseases," Blake explained.

As a next step, "we will go through an enrichment process which will be adding information to a biomarker object that will be studied and integrated with information on everything from pathways to mutations ... anything that maybe informative and valuable to our customers who are working on finding a cure for cancer."

The team plans to use the BioXM knowledge management system, developed by Sophic partner Biomax, to mine, curate, and annotate biomarker information from biomedical databases and text sources.

BioXM combines data on genes, proteins, compounds, treatments, and diseases and presents the information "graphically," thus providing an "informative map" of the components that are involved in disease, Blake said.

In addition, Blake said that a team of scientists led by Michael Liang, Sophic's principal investigator on the project, will also perform "manual curation, read papers, [and] annotate the genes and the biomarker panels so that the information is well qualified."

In the past, the Sophic/Biomax team has worked with the NCI to develop databases like the Cancer Gene Index, a collection of information on 6,955 human genes linked to cancer. For this project, researchers used Biomax's BioLT Linguistics software to automatically analyze more than 18 million abstracts in the Medline database.

Later, NCI awarded Sophic and Biomax \$1.3 million to fully annotate 3,168 brain, ovarian, and lung cancer genes that were previously identified in the Cancer Gene Index project (*BI 07/23/2008*). The company provides these annotations through its Cancer Genome Atlas database.

Both the Cancer Gene Index and Cancer Genome Atlas are freely available via the company's website.

Sophic aims to release a beta version of SCan-MarK in the fall of 2011 and to launch the first full version in 2012.

"We are calling it two iterations," he said, explaining that while the beta version of the database will incorporate data from the first five cancer types, for the full release Sophic plans to incorporate data from several other cancer types.

"One of the things that is interesting... is many of the genes that are found in a disease type are found in multiple disease types," Blake said "We are going to segment those multiple disease biomarkers and that will be also be part of the construct of how we put the system together."

Sophic will offer SCan-MarK under two separate licensing schemes geared toward government agencies, academics, and not-for-profit institutions on the one hand and biotechnology and pharmaceutical companies on the other. Blake declined to provide specific details about pricing but did say that the company plans to license the software on a "term basis."

Users will be able to install the software on site or access a web-based version hosted by Sophic through a "wiki-style interface."