

**Product:** BioXM™ Knowledge Management Environment with Viscovery® Data Mining

**Applications:** biomedical and clinical knowledge management, predictive data mining and knowledge discovery

**Product contact:** info@biomax.com

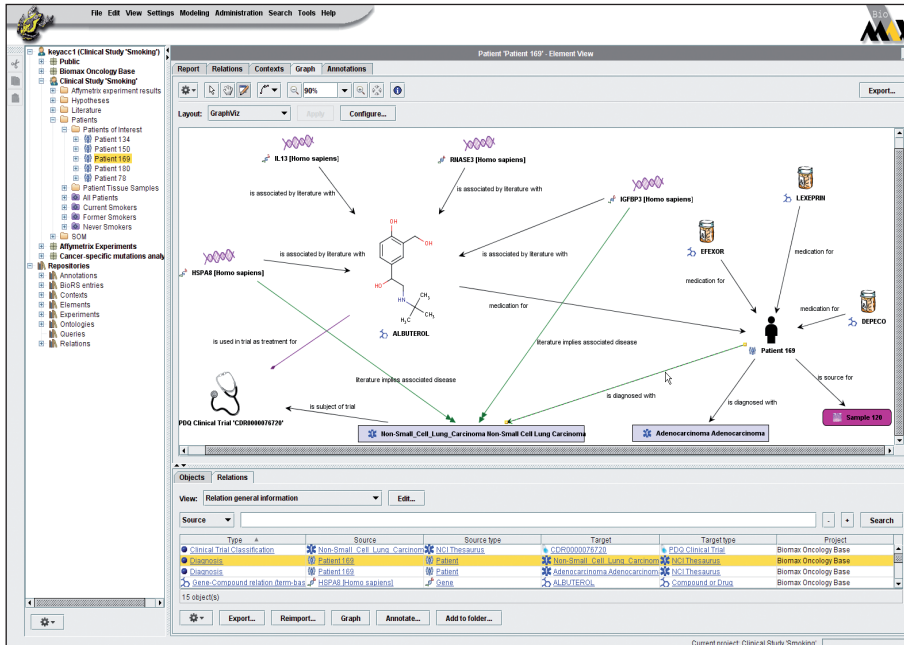


Fig. 1: BioXM visualization of relations between patients, compounds, diseases and genes

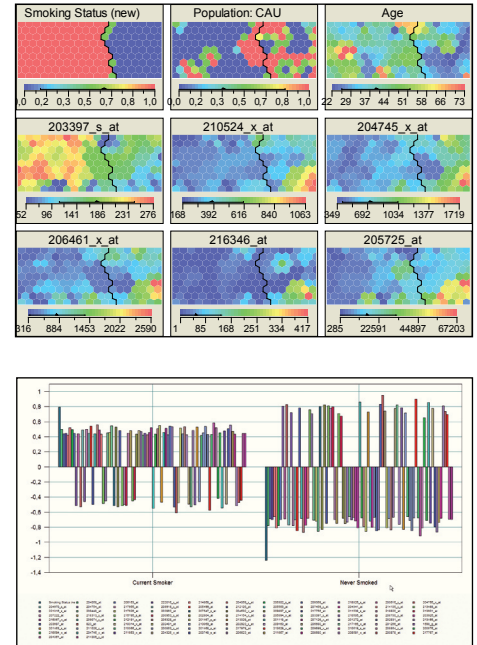


Fig. 2: Model evaluation predictive data mining

## A data mining solution for biomedical knowledge discovery

Clinical research today is faced with the challenges of translating results from advanced molecular methods into patient benefit and better understanding complex diseases such as cancer. Identification and validation of specific biomarkers with a high predictive value are key to advanced diagnostics and treatment. Indeed, the vast quantities of unstructured information generated by academic, clinical and industrial research groups are reflected in a rapidly growing body of literature; an ever-increasing number of databases and data warehouses are trying to structure and organize this information for specialized disease areas. The BioXM Knowledge Management Environment offers a comprehensive, intuitive and visual access point to these resources and puts them into context with your own data. Combining gene, disease and compound information with patient data, sample information, as well as treatment and diagnosis information becomes an easy, time-efficient task. Figure 1 demonstrates such a patient-centric view, in which terabytes of data are distilled into one picture.

Consistent management of current knowledge in the context of patient data and experimental results are the prerequisites for new data-mining approaches which support systematic discovery and prediction by exploration. Discovery and validation of biomarkers suitable for the diagnosis of specific diseases or the prediction of the outcome of therapies based on mathematical models are typical examples. Simply asked, is a specific biomarker, e.g., a set of differentially expressed genes, specifically correlated with a clinical parameter? Is this marker really capable of differentiating between the attributes I am interested in? The maps in Figure 2 provide an intuitive representation of a well-segmented information space for a specific phenotype state.

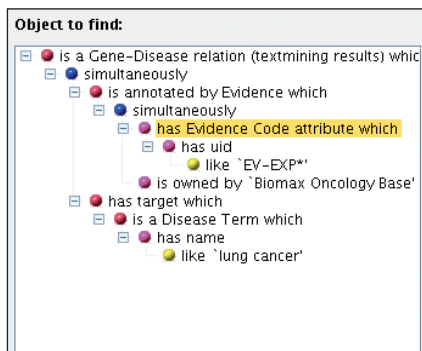
This highly visual approach, in combination with advanced algorithms, helps researchers to identify key attributes of scientific processes and disease origins and, for example, establish biomarkers for classification and diagnosis. The underlying statistical model can be used for predictive modeling, classification and forecasting of disease processes based on the key attributes identified. Using the Viscovery technology, an established biomarker can become the basis of a model classifying disease subtypes or forecasting the outcome of specific treatments. Incorporating semantic technologies and predictive data mining, the BioXM solution provides a unique research platform ideally suited to address the complexity of clinical research environments from bench to bedside.

### Biomax Informatics AG

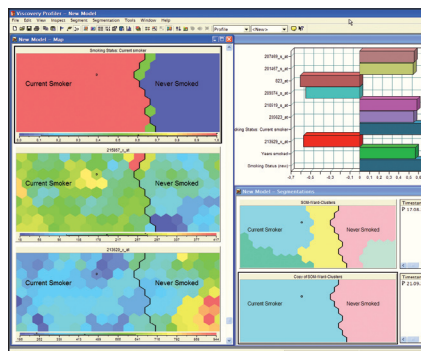
Lochhamer Str. 9, 82152 Martinsried, Germany, +49 89 895574-0

www.biomax.com/bioxm

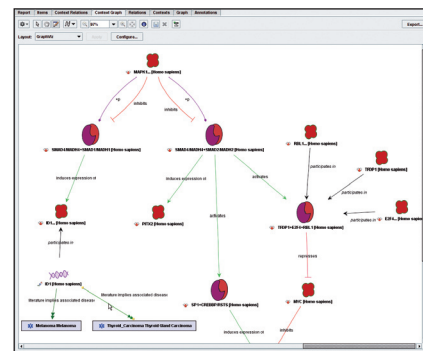




BioXM advanced query builder



Segmentation model smoker/non-smoker



Pathway contexts in the BioXM system

### Data integration using semantic technologies

The BioXM Knowledge Management Environment is designed for the aggregation and semantic modeling of scientific processes. Nearly every aspect of the BioXM system can be configured without the need to write new software or to know anything about relational database schemas. All configuration is done using the graphical user interface the BioXM client application provides.

- The semantic of a scientific area of interest (e.g., clinical research) is captured by creating a network model of related, scientifically relevant elements.
- The user can easily configure different element and relationship types, e.g., elements of type "gene" or "protein" can be linked together using a relation of type "gene regulation" or "protein-protein interaction".
- A context (a region within the larger knowledge network) allows modeling and organization of pathways.
- Experimental data (e.g., expression data tables) can be integrated as specialized BioXM objects.
- All semantic objects (e.g., relations or contexts) can be annotated.
- Entire areas of interest can be organized using ontologies.

### Viscovery predictive data-mining technology

The optional Viscovery Profiler module enables explorative knowledge discovery that goes beyond simple analysis. A highly visual approach combined with advanced algorithms helps researchers to identify key attributes of scientific processes and disease origins and, e.g., establish biomarkers for classification and diagnosis.

The Viscovery Predictor component allows predictive modeling, classification and forecasting of disease processes based on the key attributes identified with the Profiler technology. Using Predictor, a biomarker established with Profiler would become the basis of a model classifying disease subtypes or forecasting the outcome of specific treatments.

- Optimized workflow to generate multivariate statistical models
- User-defined prioritization of attributes of interest to establish biomarkers
- Model visualization for each attribute using two-dimensional projections based on self-organizing maps
- Automatic clustering or manual segmentation of target groups
- Automatic classification and prediction based on established markers

### An open, modular platform for semantic data integration, text analysis and data mining

The BioXM platform is a client-server based enterprise system. The primary user interface provides a comfortable, visual and highly interactive user experience. The BioXM server has modular, open architecture, thus ensuring expandability and maintainability. Multiple hardware platforms and operating systems are supported.

The functionality of the BioXM system can be extended by adding other optional modules. These include the virtual integration of public and proprietary scientific databases using the BioRS™ Integration and Retrieval System and text-mining capabilities provided by the BioLT™ Literature Mining Tool. The BioLT tool combines biological and medical term dictionaries with powerful free-text querying capabilities, which allow easy building of special-interest knowledge bases.

To further ensure interoperability, applications can be built on top of the BioXM server using a Web Services API. An external query mechanism allows users to extend the BioXM search engine with external tools, which provide, for example, connectivity with statistic tools or other analysis software.

*Biomax, BioLT, BioRS and BioXM are registered trademarks of Biomax Informatics AG in Germany and other countries. Viscovery is a registered trademark of Viscovery software GmbH. Java is a trademark of Sun Microsystems, Inc. The BioXM software was developed in collaboration with RingRows GmbH, Martinsried. Registered names, trademarks, etc., used in this document, even when not specifically marked as such, are not to be considered unprotected by law. BIOXMUCV0907*

