

Biomax BioXM™ information and knowledge management

The underlying idea of the BioXM™ Master Suite is to allow scientists to model a disease area in relationship to current knowledge as depicted in Figure 1 to obtain an integrated view of the processes in the context of networked biological systems, including metabolic pathways, signal transduction pathways and regulation.

BioXM in silico model

The BioXM platform is a project-centered, distributed environment that provides the data security and integration of all relevant information necessary for a disease specific modeling application: experimental data, modeling data, structured annotation, information

from public or proprietary sources (implemented using the *embedded* database integration capability of the BioRS Integration and Retrieval System).

This approach supports user-defined "objects of interest", such as "gene," "protein" or "patient," that describes a scientific field. The scientist can easily access nearly any type of experimental data, analysis results or other types of information (including gel scans, micrographs) using these conceptually defined objects. Relations between various objects such as genes, proteins, and metabolites as well as patients, disease and reactants, can be generated using simple, user-defined rules as outlined in Figure 2. Further classification and structuring of information can be implemented using a generic, high-performance annotation system which allows information in practically any form to be attached to objects as well as new or existing ontologies to be used as an informational backbone.

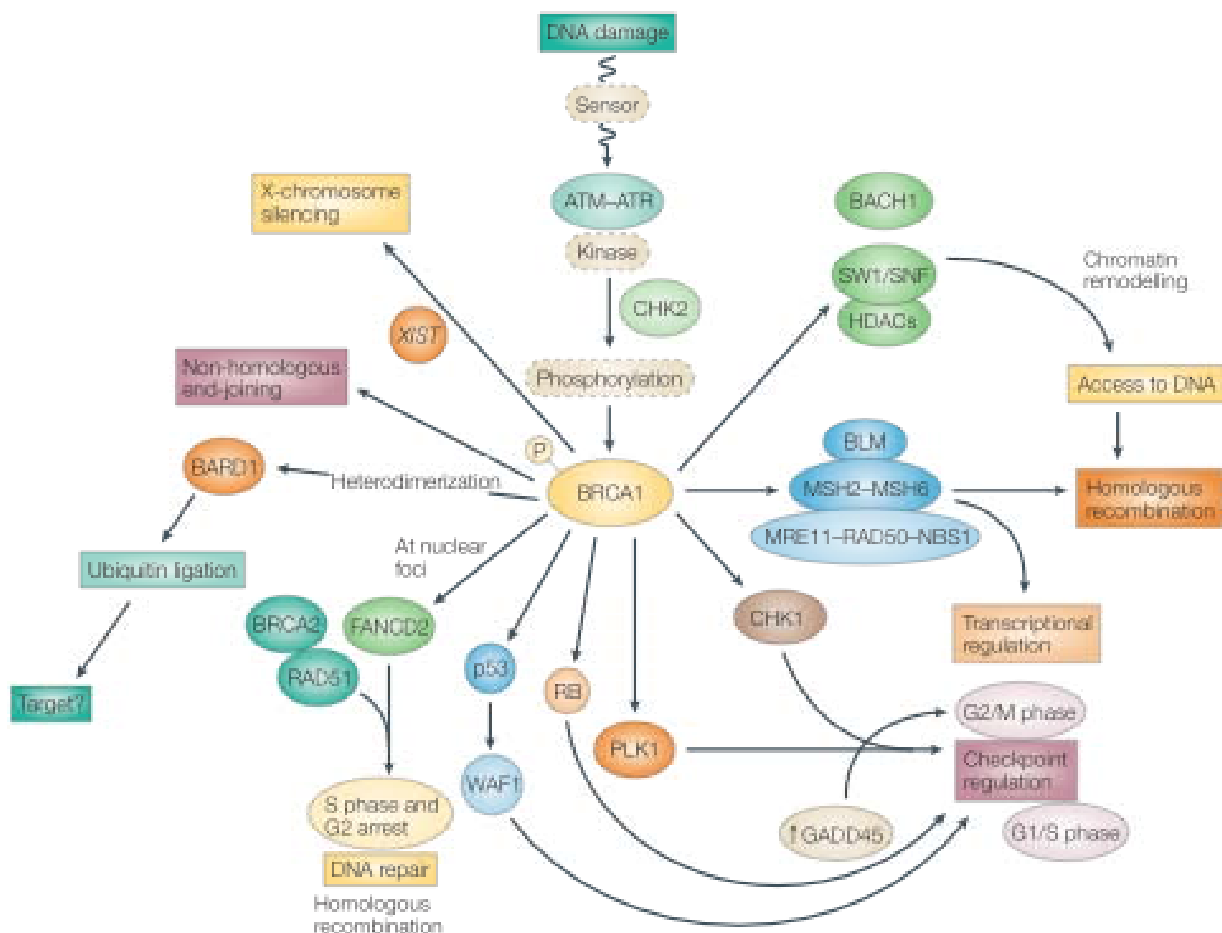


Figure 1 Narod, SA and Foulkes, WD (2004) BRCA1 and BRCA2: 1994 and beyond. *Nature Reviews Cancer* 4, 665-676.



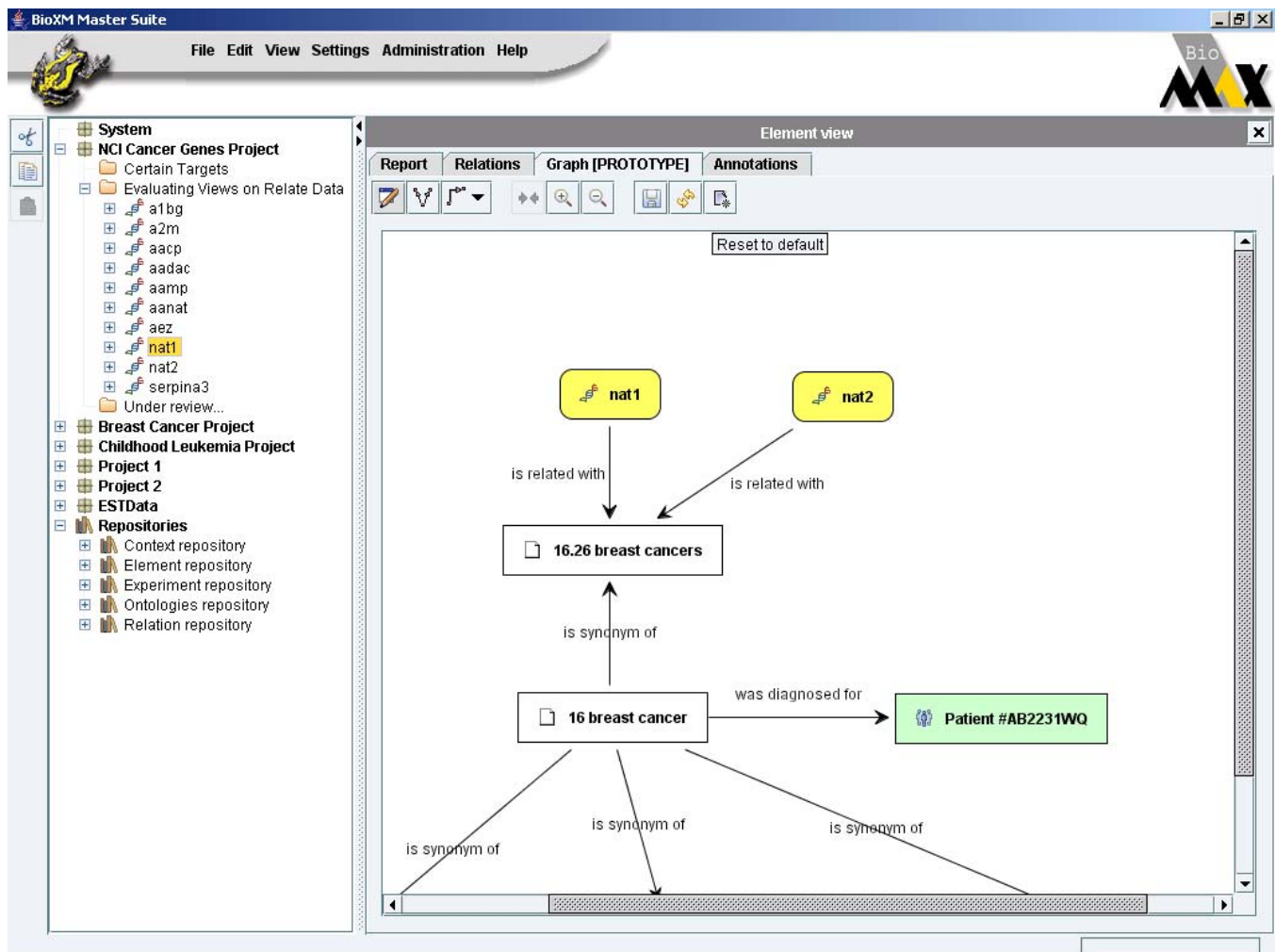


Figure 2 Screenshot of a breast cancer sub-model depicted from the NCI cancer gene index project which has been extended to include patient information

BioXM Information management concepts

The BioXM Master Suite uses existing concepts from the areas of knowledge management for the aggregation and semantic modeling of biological processes. In particular, the following are included:

- A defined problem area (e.g., a disease area) is semantically modeled as a network of related objects (elements).
- The elements and their relationships are classified; i.e., elements of type "gene" or "protein" can be linked together using a relation of type "gene regulation" or "protein–protein interaction."
- Element and relation types can be defined by the user.
- A context is a region within the knowledge network and is used for modeling knowledge. A context is a semantic object and, like elements, can be linked to an element and other semantic objects.
- Additional meta-information can be attached to each semantic object (such as element or relation) in the BioXM Master Suite in the form of annotation.
- Annotation offers a mechanism for attaching information and comments to objects without altering the object within the context of the model, allowing the annotation to become an integral part of the object.
- A form-based approach supports the "modeling" of related information.

- Combination of different forms (hierarchical organization) enables complex data models (e.g., MIAME and PEDRo).
- Objects can share annotation and, in this way, be used to imply relationships between objects.
- Entire problem areas can be conceptualized and organized through the use of ontologies.
- Controlled vocabularies (such as value lists, thesauri, ontologies) are supported.
- Experimental data (e.g., data table from experimental measurements) can be organized within the BioXM Master Suite as objects.
- The BioXM Master Suite and the Biomax BioRS Integration and Retrieval System seamlessly bind a classical integration platform for biological databanks and extends it with semantic components

Technical description

The BioXM Master Suite is a client-server software system having multiple tiers.

A client communicates using the HTTP/S protocol with a Web server, which communicates with the BioXM server using the Common Object Request Broker Architecture (CORBA). The BioRS server can be

accessed using either CORBA or HTTP/S. Additional servers for Resource and User Management (RUM) and relational database management systems are accessed using the local, server-side network. All servers can be hosted on different computers. The BioXM Server has a highly modular architecture. The goal of this architecture is to ensure expendability and maintainability.

A module represents the simplest possible technical and functional unit that, as a rule, a GUI, server (Business Logic) or database component comprises (client, server or database tier). Ideally a module represents a service that is integrated with other modules only using the CORBA or other derived Java interface and ideally only in a single tier (e.g., only the GUI tier). Figure 3 provides an overview of the modular open architecture.

The primary user client is a Java-Swing graphical user interface (GUI). In addition, BioXM can generate Web-based reports, which may be accessed with any Web browser. Therefore, the GUI offers all the interactive function that is expected from an application software (e.g., drag-and-drop, copy and paste using a clipboard, printing and access to the file system). In addition, the familiar functionality and concept that a Web browser provides (e.g., links) are also available.

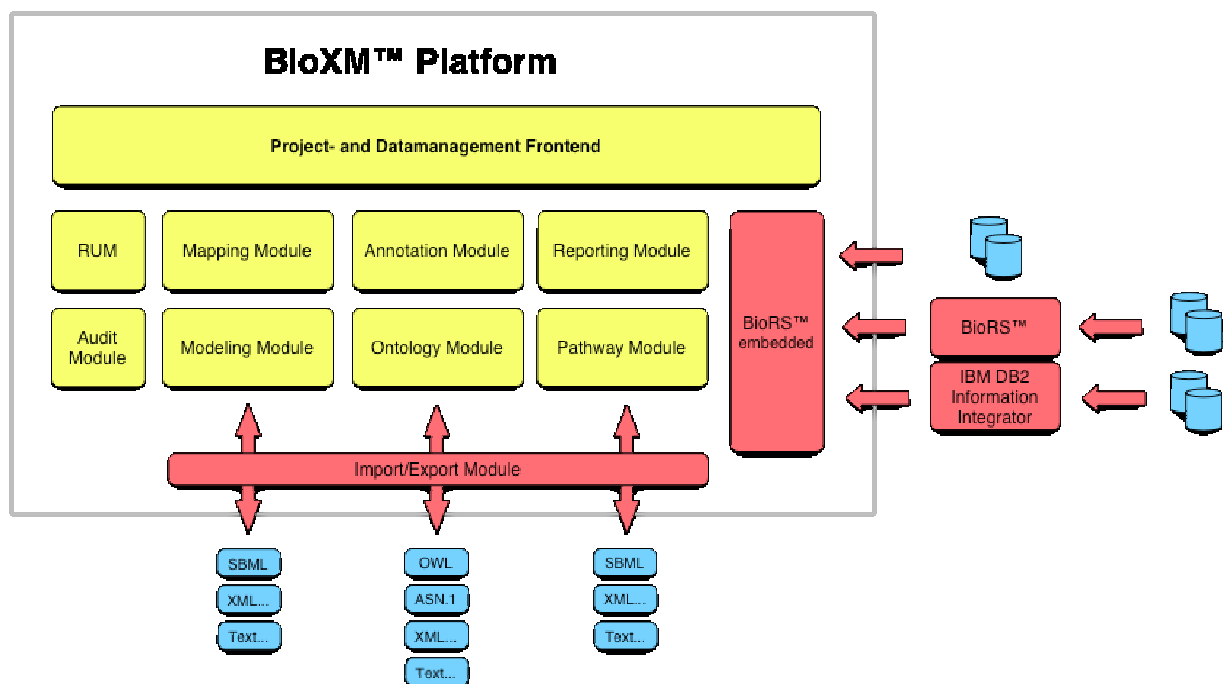


Figure 3 BioXM overview of modules and interoperability with middleware and data access components

