Knowledge networks of biological and medical data

An exhaustive and flexible solution to model life sciences domains

Dr. Sascha Losko, Dr. Karsten Wenger, Dr. Wenzel Kalus, Dr. Andrea Ramge, Dr. Jens Wiehler, Dr. Klaus Heumann

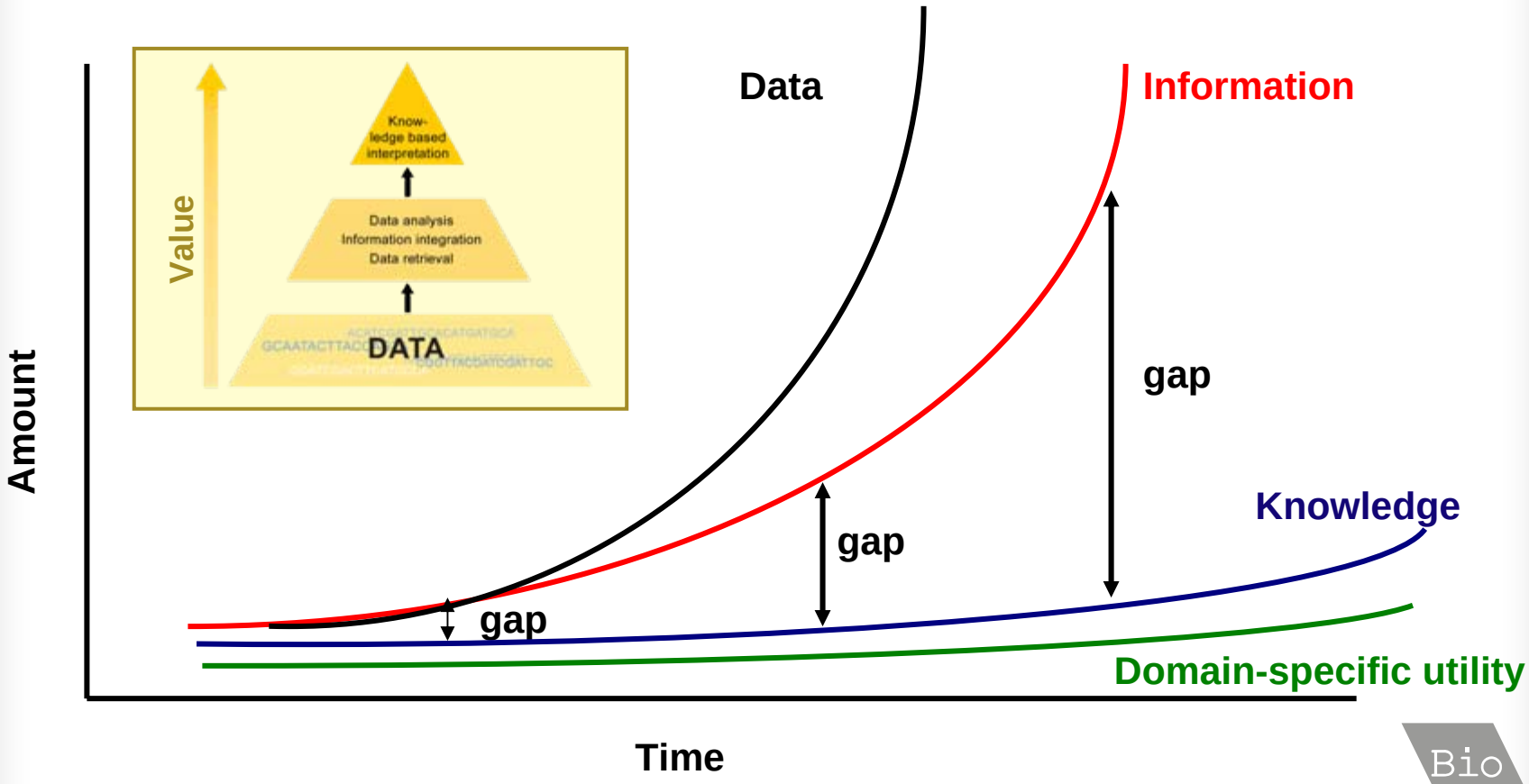
Biomax Informatics AG
provides novel solutions
for better decision making
and knowledge management

http://www.biomax.com

Overview

- Motivation and Concepts
- BioXM™ Knowledge Management Environment –
  a System for Domain Modeling and Semantic Integration
- Applications in e.g. the Oncology domain
- Textmining-based Knowledge Capturing
- Knowledge Presentation, Mining and Processing
- Conclusion

Knowledge Gap in Life Sciences



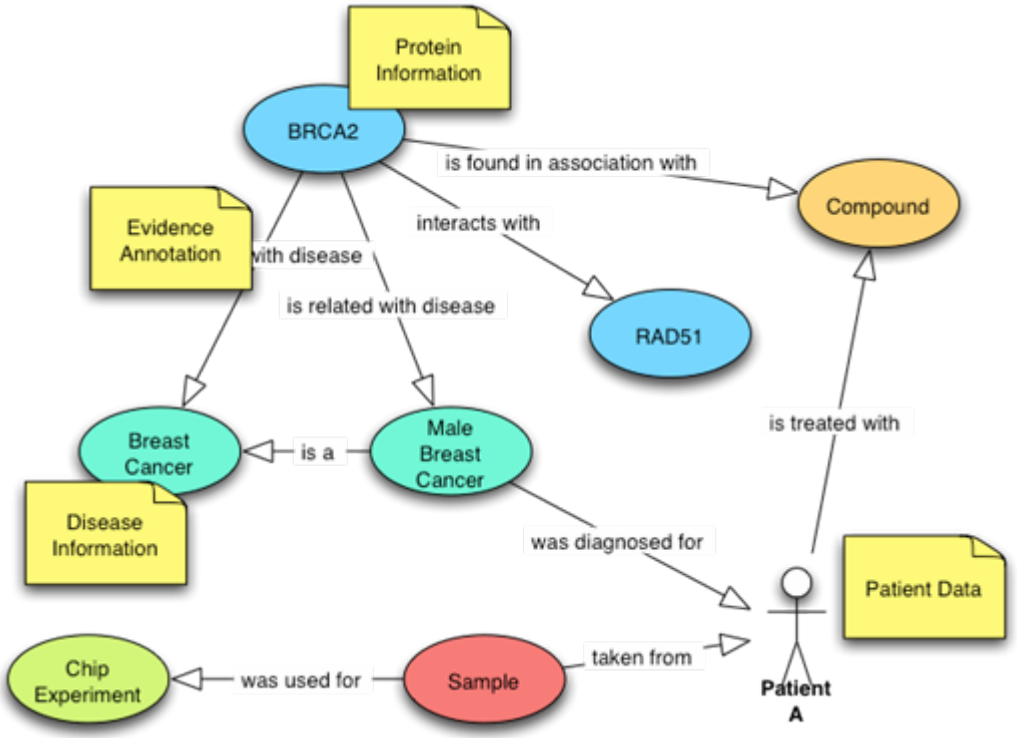
Biomax Informatics AG

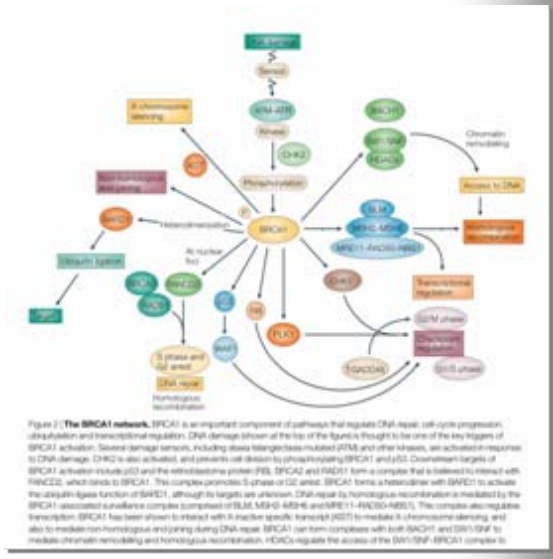
A need for software supporting knowledge management in life science

Application: How to address key questions in oncology?

- Which genes are described:
  - in association with a specific cancer type?
  - by experimental evidence?
  - to be upregulated?
- Which compounds are described
  - to inhibit a gene?
  - in context with which a cancer type?
- Which cancer types are described
  - in association with certain compounds?
  - in context of cell line assay of a target gene?
- What is the mouse ortholog of a cancer gene? Do they share a specific domain?

BioXM Technology Concept - *In-Silico* Knowledge Representation





Narod, S.A. and Foulkes, W.D. (2004) BRCA1 and BRCA2: 1994 and beyond. Nature Reviews Cancer, 4, 665-676.

- Versatile and semantically rich network representation of biomedical knowledge which is flexible and open to accommodate any type of entities and metadata
- The knowledge network is the one-stop-shop for all relevant resources: “Knowledge Inventory”

Biomax Informatics AG

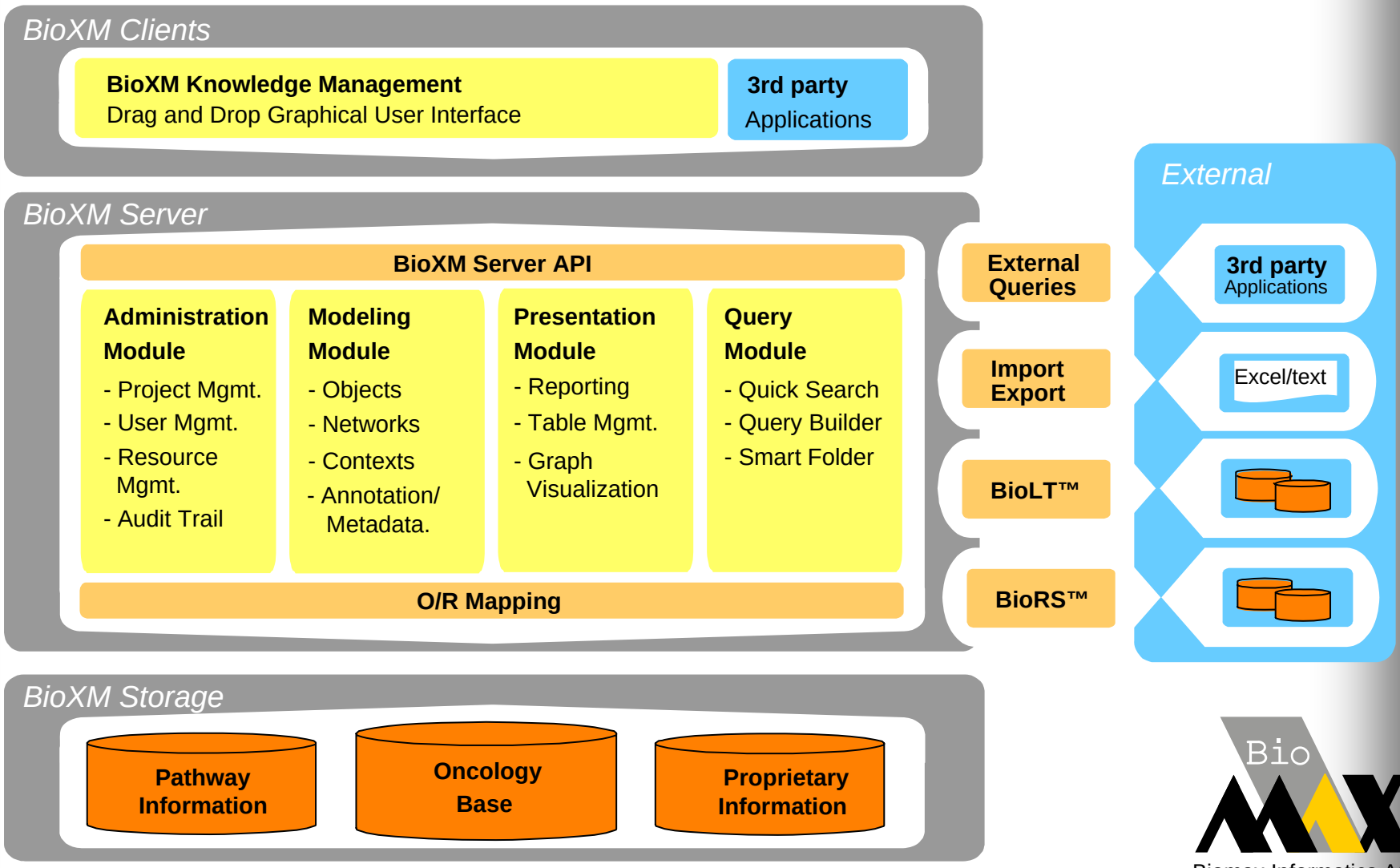
BioXM Knowledge Management Key Features

The BioXM™ platform is designed to be configured to support diverse types of scientific and biomedical knowledge management applications:

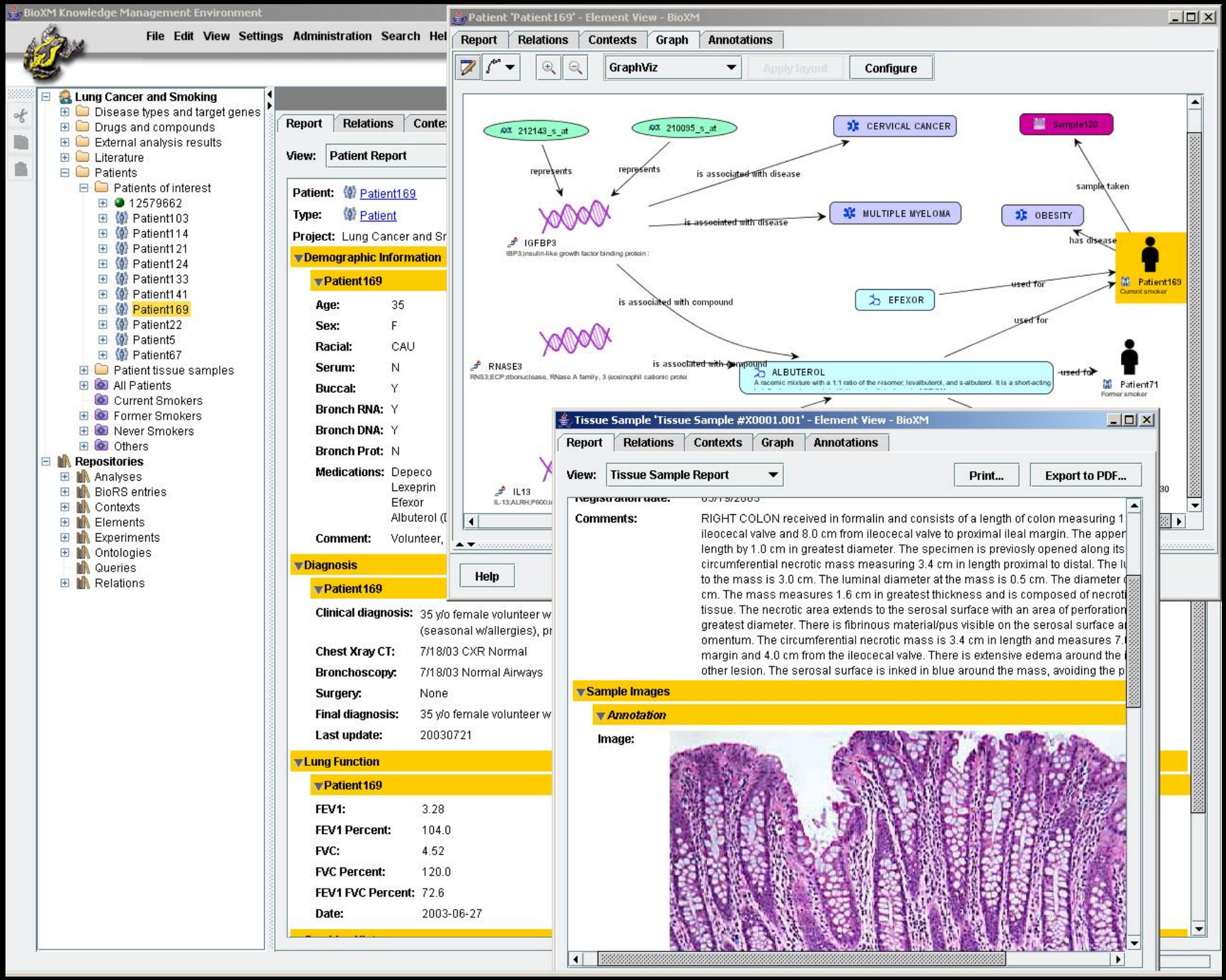
- Connects and visualizes data, information and knowledge
- Enables full data integration for discovering novel relationships and patterns in biological networks
- Query as you think and work
- On-the-fly building of new data connections and networks
- Enables mapping of proprietary knowledge on top of public ontologies
- Maintains full audit trail
- Maximum flexibility without additional programming
- Supports interoperability and standardized interfaces
- Operates on multiple relational database systems (e.g. Oracle)

Biomax Informatics AG

BioXM Technology Platform



Biomax Informatics AG


BioXM Knowledge Management Environment
File Edit View Settings Administration Search Hel
Lung Cancer and Smoking
Disease types and target genes
Drugs and compounds
External analysis results
Literature
Patients
Patients of interest
12579662
Patient103
Patient114
Patient121
Patient124
Patient133
Patient141
Patient169
Patient22
Patient5
Patient67
Patient tissue samples
All Patients
Current Smokers
Former Smokers
Never Smokers
Others
Repositories
Analyses
BioRS entries
Contexts
Elements
Experiments
Ontologies
Queries
Relations
Report
Relations
Conte
View: Patient Report
Patient: Patient169
Type: Patient
Project: Lung Cancer and Sr
Demographic Information
Patient169
Age: 35
Sex: F
Racial: CAU
Serum: N
Buccal: Y
Bronch RNA: Y
Bronch DNA: Y
Bronch Prot: N
Medications: Depeco
Lexeprin
Efexor
Albuterol (I
Comment: Volunteer,
Diagnosis
Patient169
Clinical diagnosis: 35 y/o female volunteer w
(seasonal w/allergies), pr
Chest Xray CT: 7/18/03 CXR Normal
Bronchoscopy: 7/18/03 Normal Airways
Surgery: None
Final diagnosis: 35 y/o female volunteer w
Last update: 20030721
Lung Function
Patient169
FEV1: 3.28
FEV1 Percent: 104.0
FVC: 4.52
FVC Percent: 120.0
FEV1 FVC Percent: 72.6
Date: 2003-06-27
Patient 'Patient169' - Element View - BioXM
Report
Relations
Contexts
Graph
Annotations
GraphViz
Apply layout
Configure
212143_s_at
210095_s_at
represents
represents
is associated with disease
CERVICAL CANCER
Sample120
sample taken
is associated with disease
MULTIPLE MYELOMA
OBESITY
has disease
IGFBP3
IBP3;insulin-like growth factor binding protein
Patient169
Current smoker
is associated with compound
EFEXOR
used for
used for
RNASE3
RNS3;ECP;ribonuclease, RNase A family, 3 (eosinophil cationic protei
is associated with compound
ALBUTEROL
A racemic mixture with a 1:1 ratio of the r-isomer, levalbuterol, and s-albuterol. It is a short-acting
used for
Patient71
Former smoker
IL13
IL-13;ALRH;P600
Help
Tissue Sample 'Tissue Sample #X0001.001' - Element View - BioXM
Report
Relations
Contexts
Graph
Annotations
View: Tissue Sample Report
Print...
Export to PDF...
Comments: RIGHT COLON received in formalin and consists of a length of colon measuring 1
ileocecal valve and 8.0 cm from ileocecal valve to proximal ileal margin. The apper
length by 1.0 cm in greatest diameter. The specimen is previosly opened along its
circumferential necrotic mass measuring 3.4 cm in length proximal to distal. The l
to the mass is 3.0 cm. The luminal diameter at the mass is 0.5 cm. The diameter
cm. The mass measures 1.6 cm in greatest thickness and is composed of necrot
tissue. The necrotic area extends to the serosal surface with an area of perforation
greatest diameter. There is fibrinous material/pus visible on the serosal surface a
omentum. The circumferential necrotic mass is 3.4 cm in length and measures 7.
margin and 4.0 cm from the ileocecal valve. There is extensive edema around the
other lesion. The serosal surface is inked in blue around the mass, avoiding the p
Sample Images
Annotation
Image:

Biomax Oncology Base

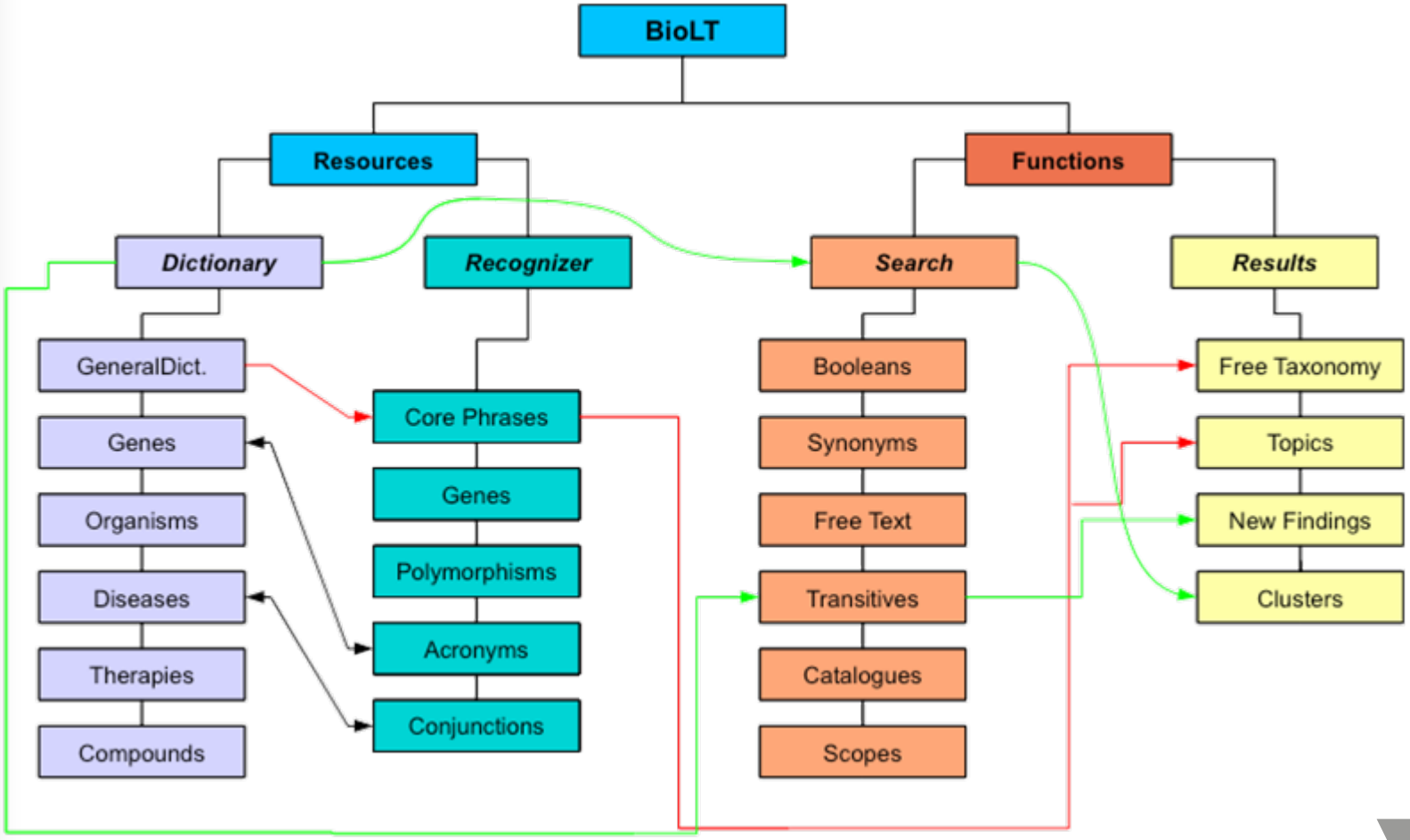
Includes the NCI *Cancer Gene Index* *

The NCI *Cancer Gene Index* is a database of associations between **genes** and **diseases** and genes and **drug compounds** derived from the biomedical literature as a single source to help cancer researchers to accelerate the search for novel cancer cures.

* In 2004 Biomax and Sophic Systems Alliance Inc. have teamed with the NCI to develop the *Cancer Gene Index*

Bio MAX
Biomax Informatics AG

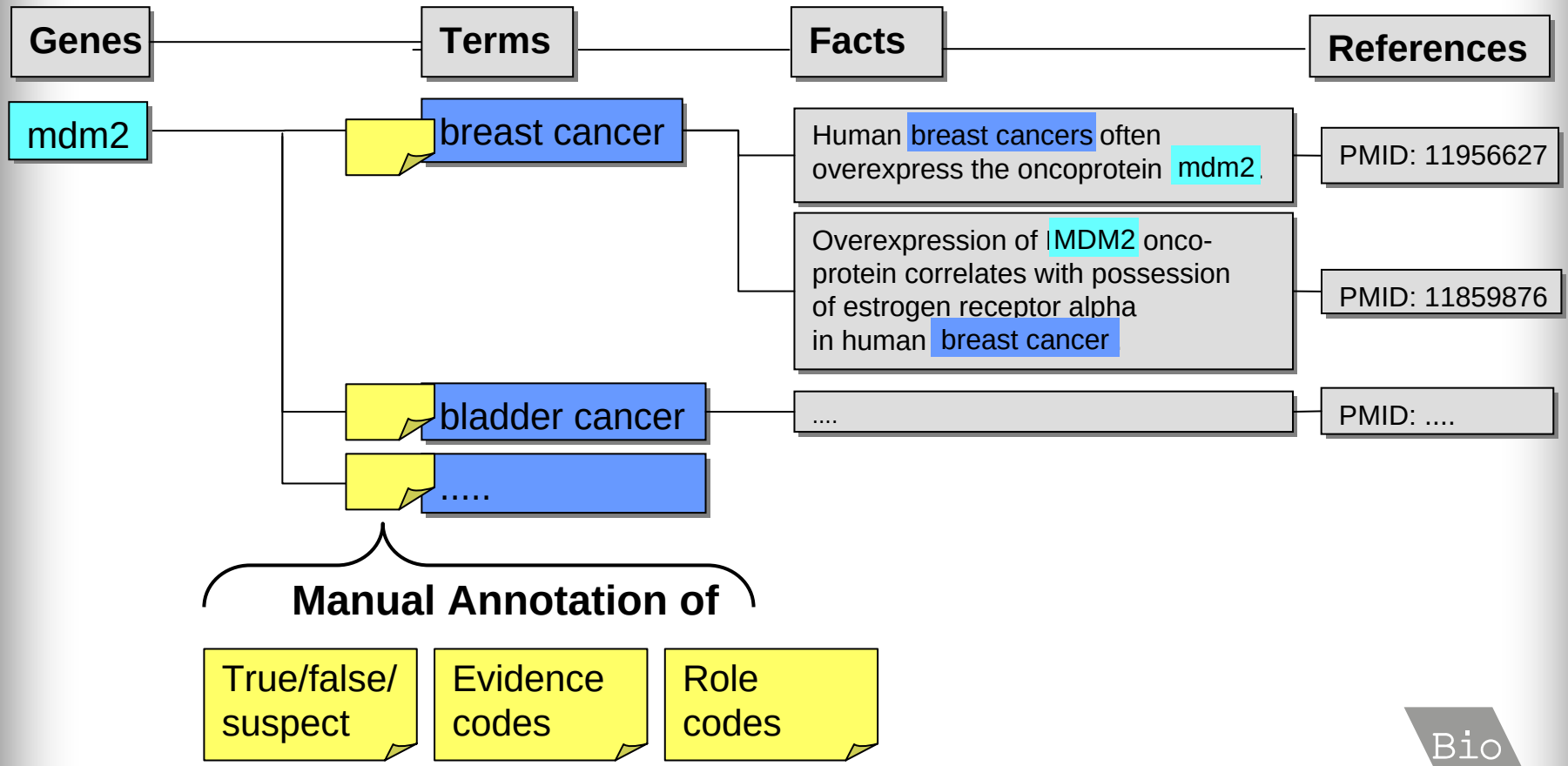
Generating the NCI Cancer Gene Index
1st Step: BioLT Textmining Engine



- Selection and configuration of the engine components allows balancing precision and recall
- To generate the NCI Cancer Gene Index, recall was optimized

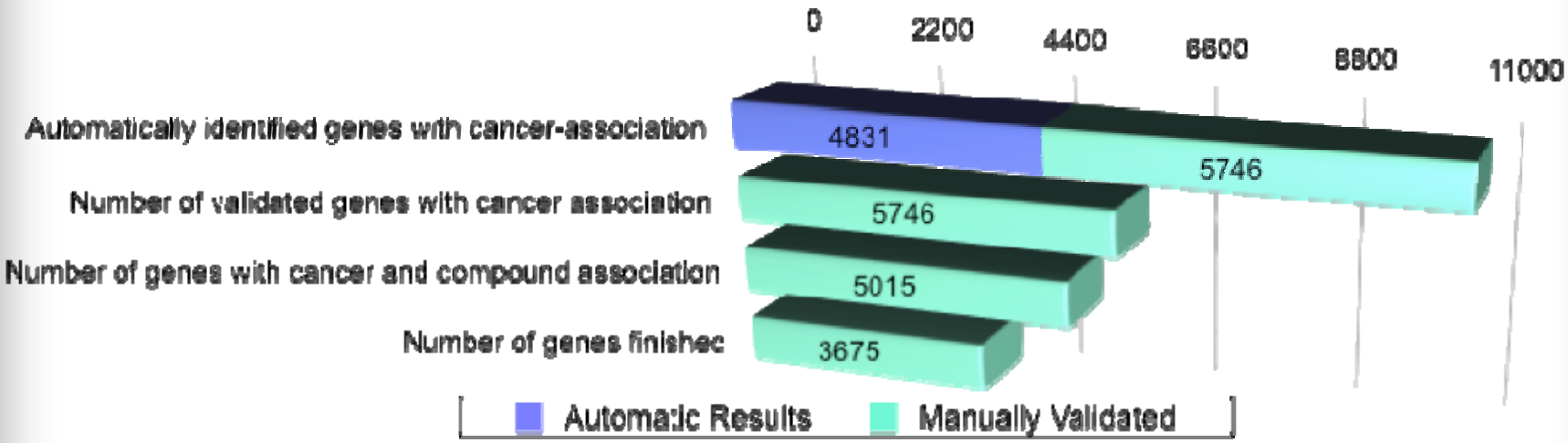
Biomax Informatics AG

2nd Step: Manual Curation



Biomax Informatics AG

Project Status



About 5,800 manually validated, “true” cancer genes (out of ~10,500 candidates)

- For **5,746 cancer genes**, **~20,000 cancer terms** and **~5,000 compound terms** have been found to be associated
- For each gene all Gene-Disease and Gene-Compound relations have been verified by experts and annotated.
- Gene-Disease specific annotations include e.g. biomarker, gene/protein expression in disease, cell line information, therapeutic relevance.
- Gene-Compound specific annotations include e.g. influence on expression, resistance, binding, transport.
- Terms have been mapped back to the “NCI Thesaurus” ontology

Biomax Informatics AG

Evidence-based classification of identified Relations

Evidence

- In average, ~316 disease-related sentences and ~380 compound-related sentences are found for each gene
- About 400,000 abstracts and ~1,370,000 sentences have been manually reviewed so far

Relations are manually classified by ontology-based codes for Evidence-type, relation roles and role details

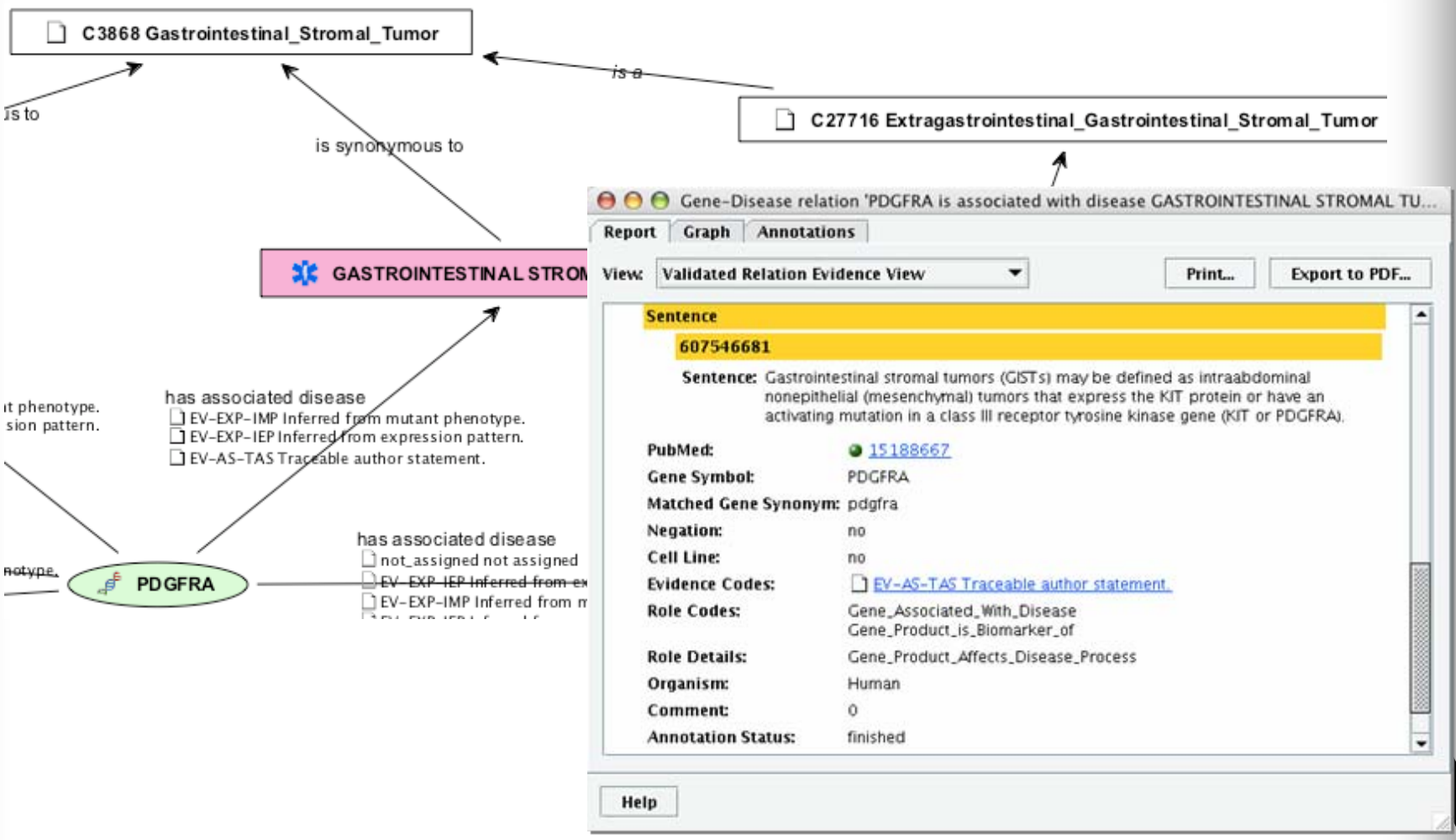
- More than 50 different codes for describing Gene-Disease relations.
- More than 40 different codes for describing Gene-Compound relations

Example:

Evidence Code	Description	Assignments	Classified Relations
EV-EXP	Inferred from experiment.	70506	34968
EV-AS	Author statement.	54135	22175
EV-COMP	Inferred from computational analysis.	426	356
EV-IC	Inferred by curator.	120	118

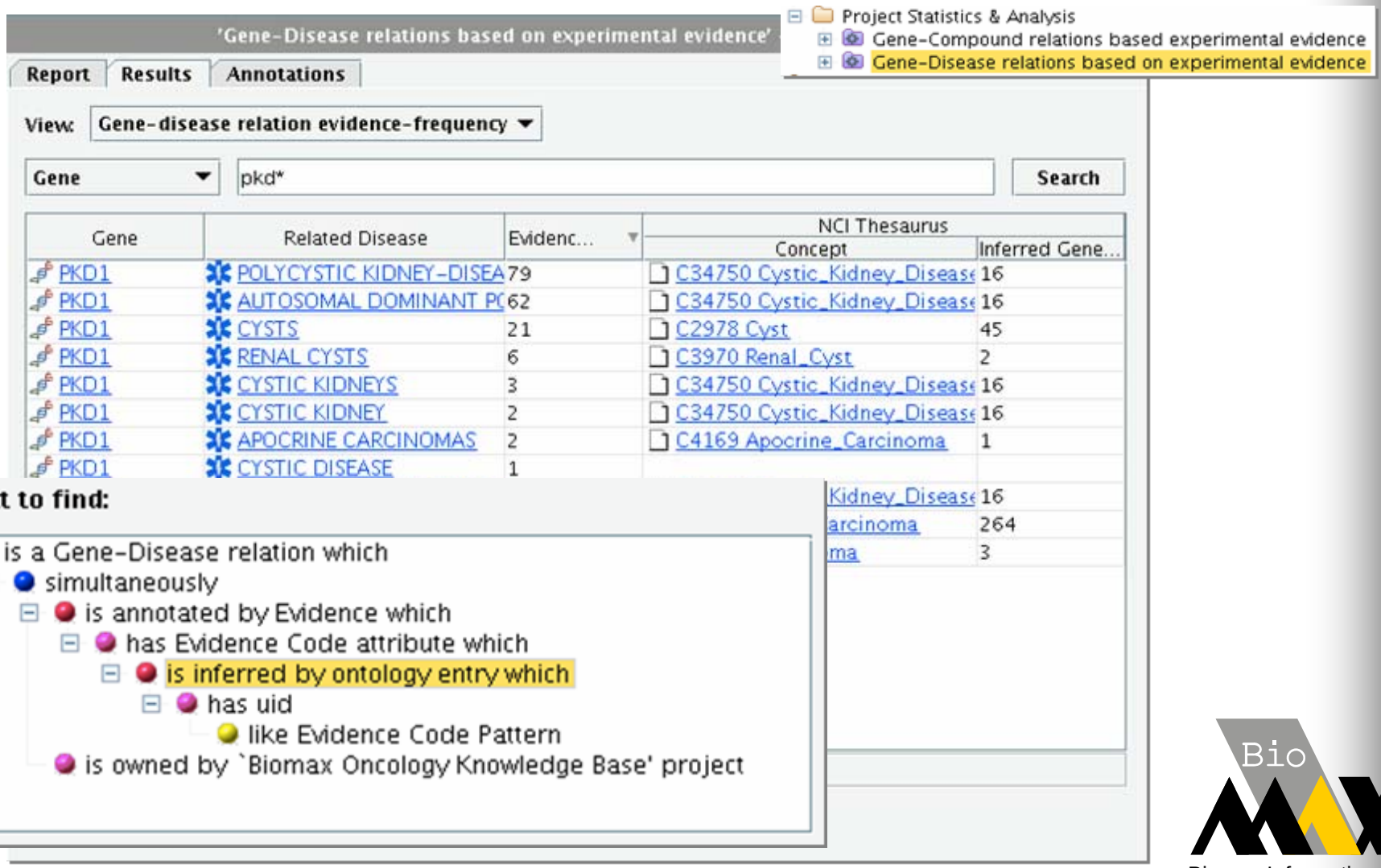
Evidence concepts (only top level shown) from Evidence Ontology (Karp et al.)

BioXM – Visualization and Editing of e.g. Textmining Results



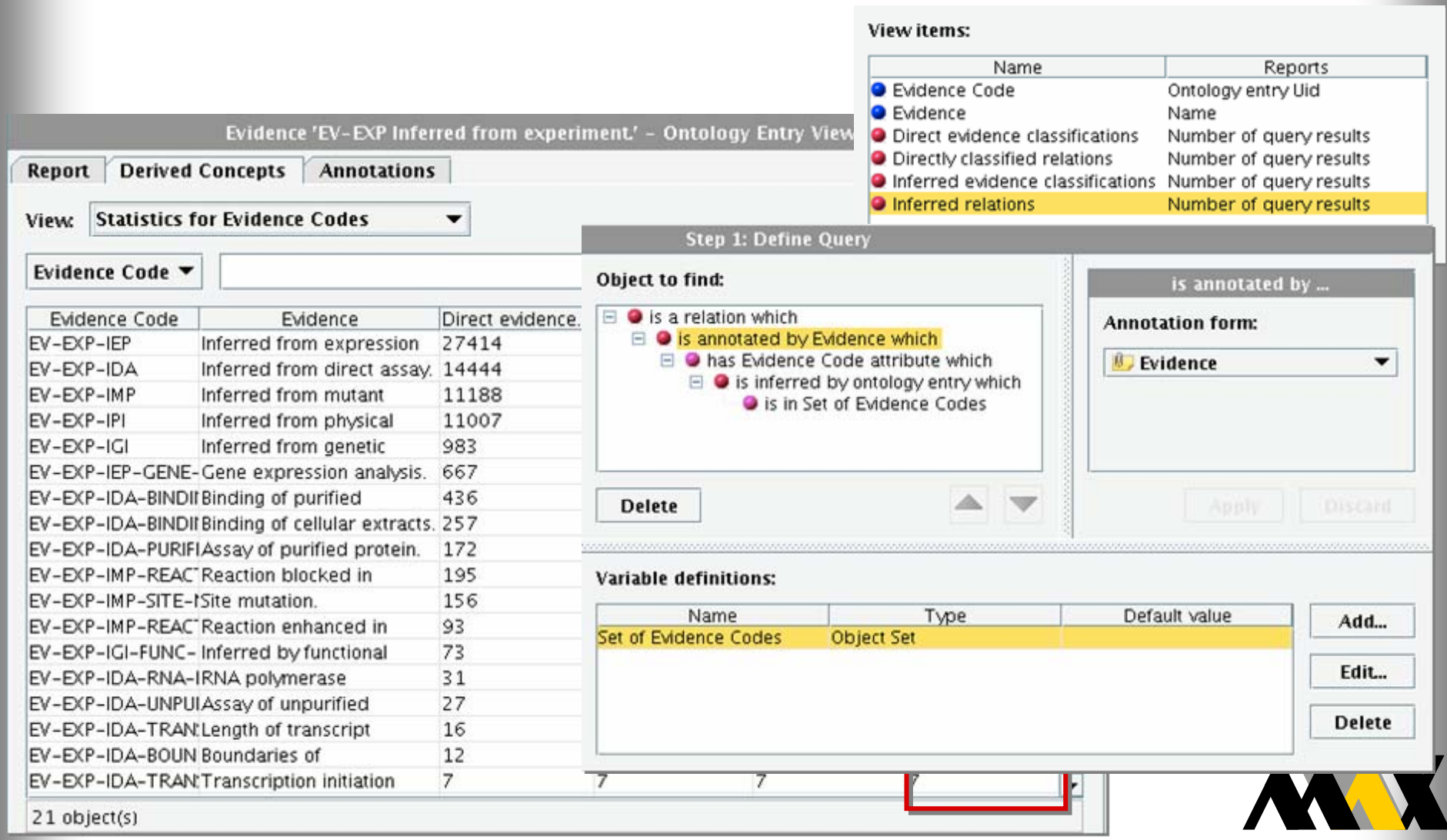
BioXM – Querying the Oncology Base

"Find genes experimentally associated with specific cancer types"



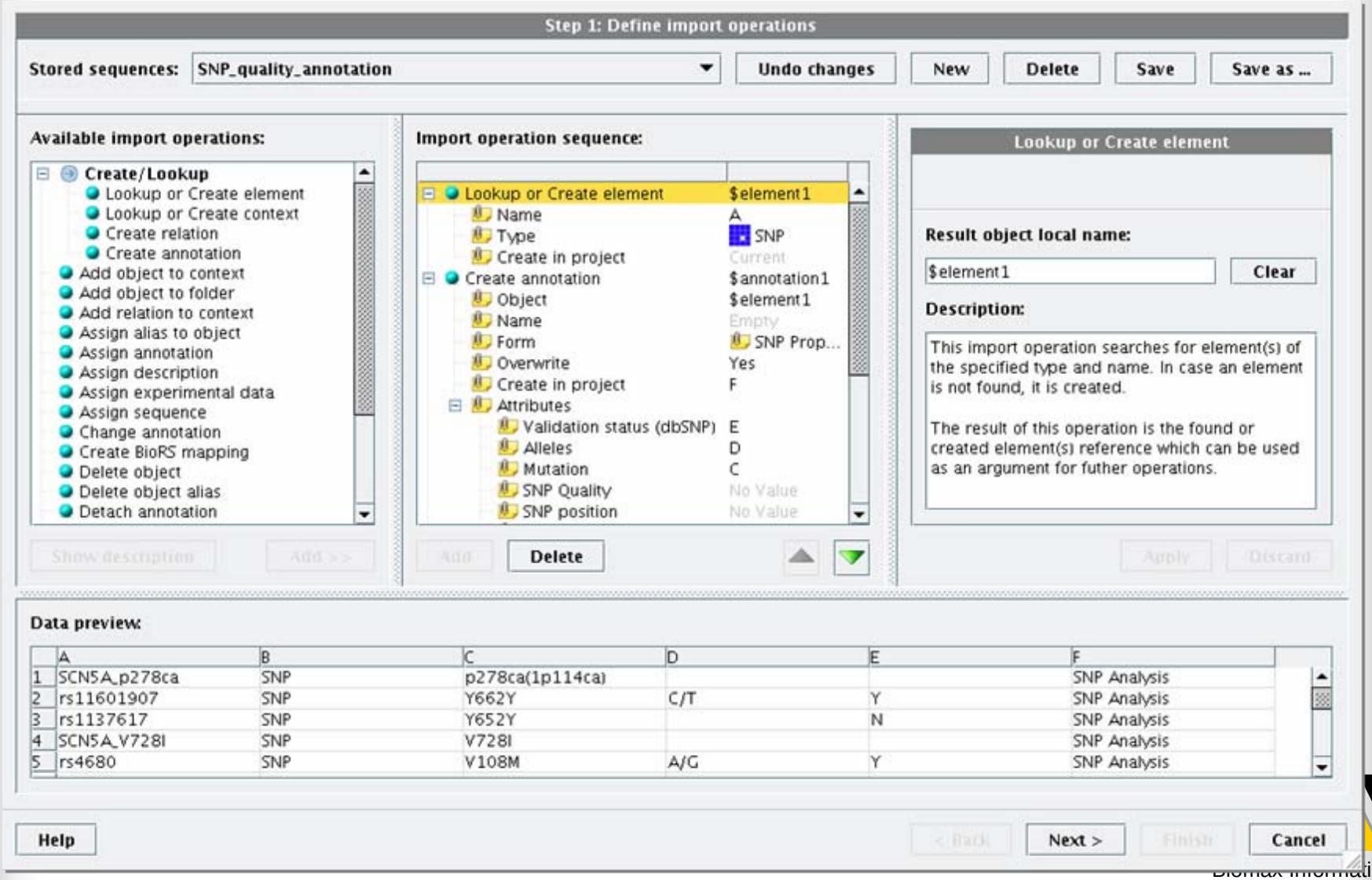
Biomax Informatics AG

BioXM – Flexible report generation with “in-view” analysis



Biomax Informatics AG

BioXM – Table-driven Knowledge Processing



Conclusion

Text mining to find all current cancer genes to establish an oncology knowledge base

NCI Project
Gene-Disease-Compounds

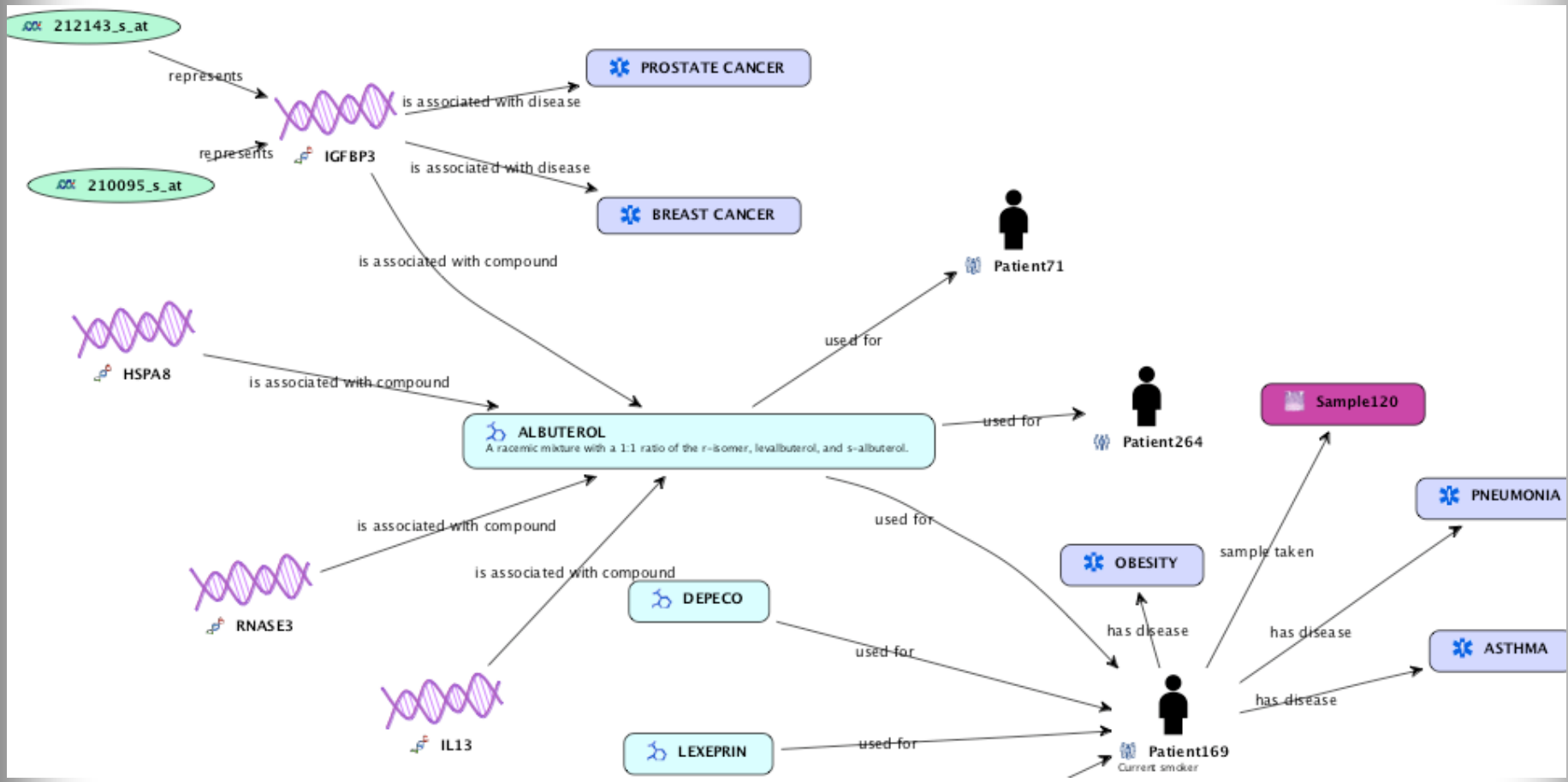
BioXM
Knowledge Management

Infrastructure for exploiting and leveraging the knowledge

Applications

- **Cancer**
- **Other Complex Diseases**
- **Research Hypothesis**
- **Validation**
- **Pipeline & Knowledge Mgmt**
- **Diagnosis**
- **Treatment**

Thank you!



Info: http://www.biomax.com/bioxm

Biomax Informatics AG