

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

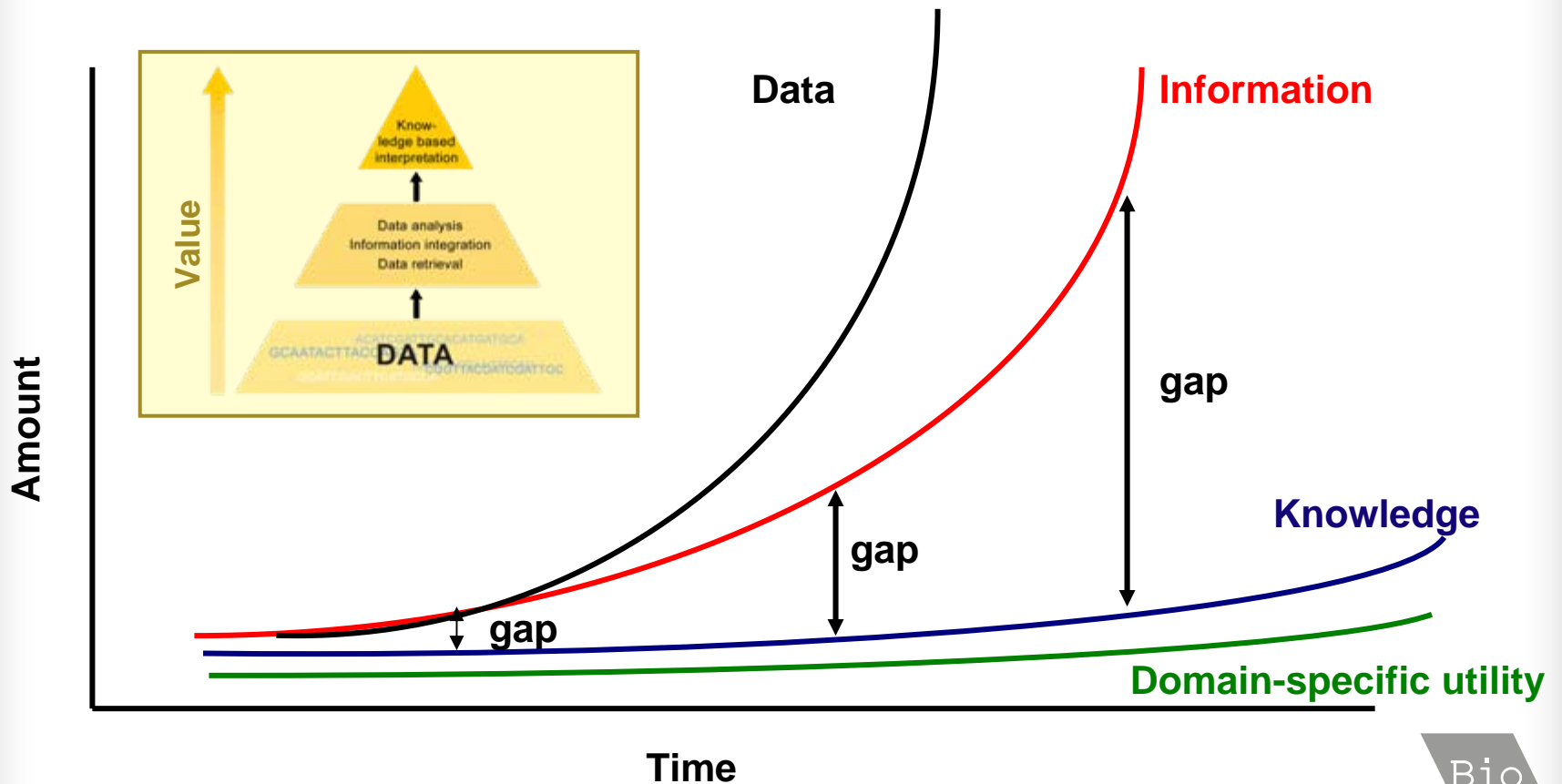
Biomax Informatics AG
provides novel solutions
for better decision making
and knowledge management

<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



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

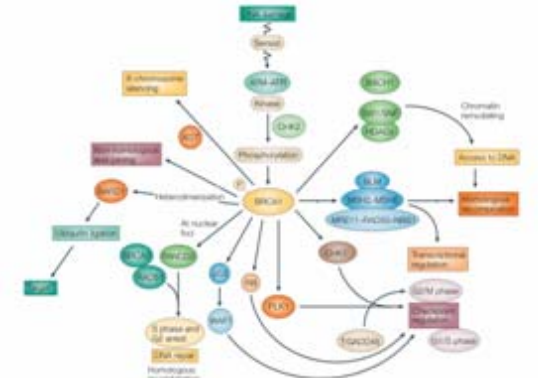
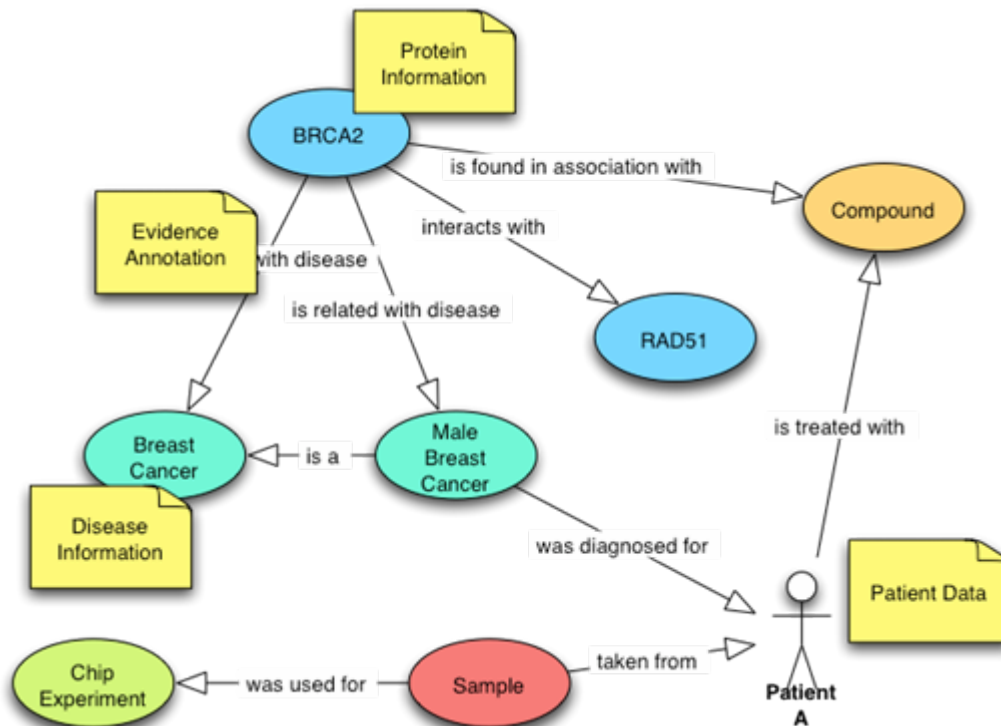


Figure 2 | The BRCA1 network. BRCA1 is an important component of pathways that regulate DNA repair, cell cycle progression, ubiquitination and transcriptional regulation. DNA damage (shown at the top of the figure) is thought to be one of the key triggers of BRCA1 activation. Several damage sensors, including ataxia telangiectasia mutated (ATM) and other kinases, are activated in response to DNA damage. ATM is also activated, and prevents cell death by phosphorylating BRCA1 and p53. Downstream targets of BRCA1 activation include p53 and the restriction endonuclease (REB1). BRCA1 and FANCD1 form a complex that is believed to interact with FANCD2, which binds to BRCA1. This complex promotes G phase to G2 arrest. BRCA1 forms a heterodimer with BRCA2 to activate the ubiquitin ligase function of BRCA1, although its targets are unknown. DNA repair by homologous recombination is mediated by the BRCA1-associated surveillance complex composed of SLM, MRE11, RAP80 and MRE11-FANCD1-MRE11. This complex also regulates transcription. BRCA1 has been shown to interact with its nuclear specific transcription factor (NBS1) to regulate X chromosome silencing, and also to mediate non-homologous end joining during DNA repair. BRCA1 can form complexes with both BRCA2 and DNA-PK to mediate chromosome remodeling and homologous recombination. mRNAs regulate the access of the DNA-PK-BRCA1 complex to

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”

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)

BioXM Technology Platform

BioXM Clients

BioXM Knowledge Management
Drag and Drop Graphical User Interface

3rd party
Applications

BioXM Server

BioXM Server API

Administration Module

- Project Mgmt.
- User Mgmt.
- Resource Mgmt.
- Audit Trail

Modeling Module

- Objects
- Networks
- Contexts
- Annotation/ Metadata.

Presentation Module

- Reporting
- Table Mgmt.
- Graph Visualization

Query Module

- Quick Search
- Query Builder
- Smart Folder

O/R Mapping

External Queries

Import Export

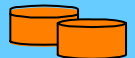
BioLT™

BioRS™

External

3rd party
Applications

Excel/text



BioXM Storage

Pathway
Information

Oncology
Base

Proprietary
Information



- 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
 - Patent 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

▼ Patient 169

Age: 35

Sex: F

Racial: CAU

Serum: N

Buccal: Y

Bronch RNA: Y

Bronch DNA: Y

Bronch Prot: N

Medications: Depeco
Lexepri
Efexor
Albuterol (I

Comment: Volunteer,

Diagnosis

▼ Patient 169

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

▼ Patient 169

FEV1: 3.28

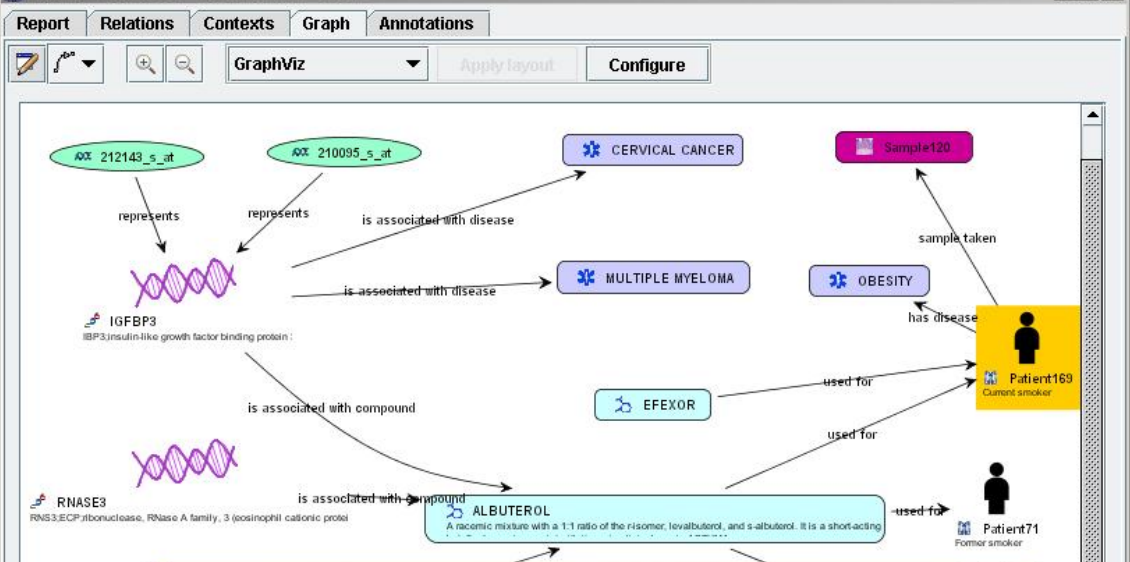
FEV1 Percent: 104.0

FVC: 4.52

FVC Percent: 120.0

FEV1 FVC Percent: 72.6

Date: 2003-06-27



Tissue Sample "Tissue Sample #X0001.001" - Element View - BioXM

Report Relations Contexts Graph Annotations

View: Tissue Sample Report

Print... Export to PDF...

Registration date: 03/19/2003

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 upper length by 1.0 cm in greatest diameter. The specimen is previously 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 necrotic 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 at omentum. The circumferential necrotic mass is 3.4 cm in length and measures 7.1 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* *

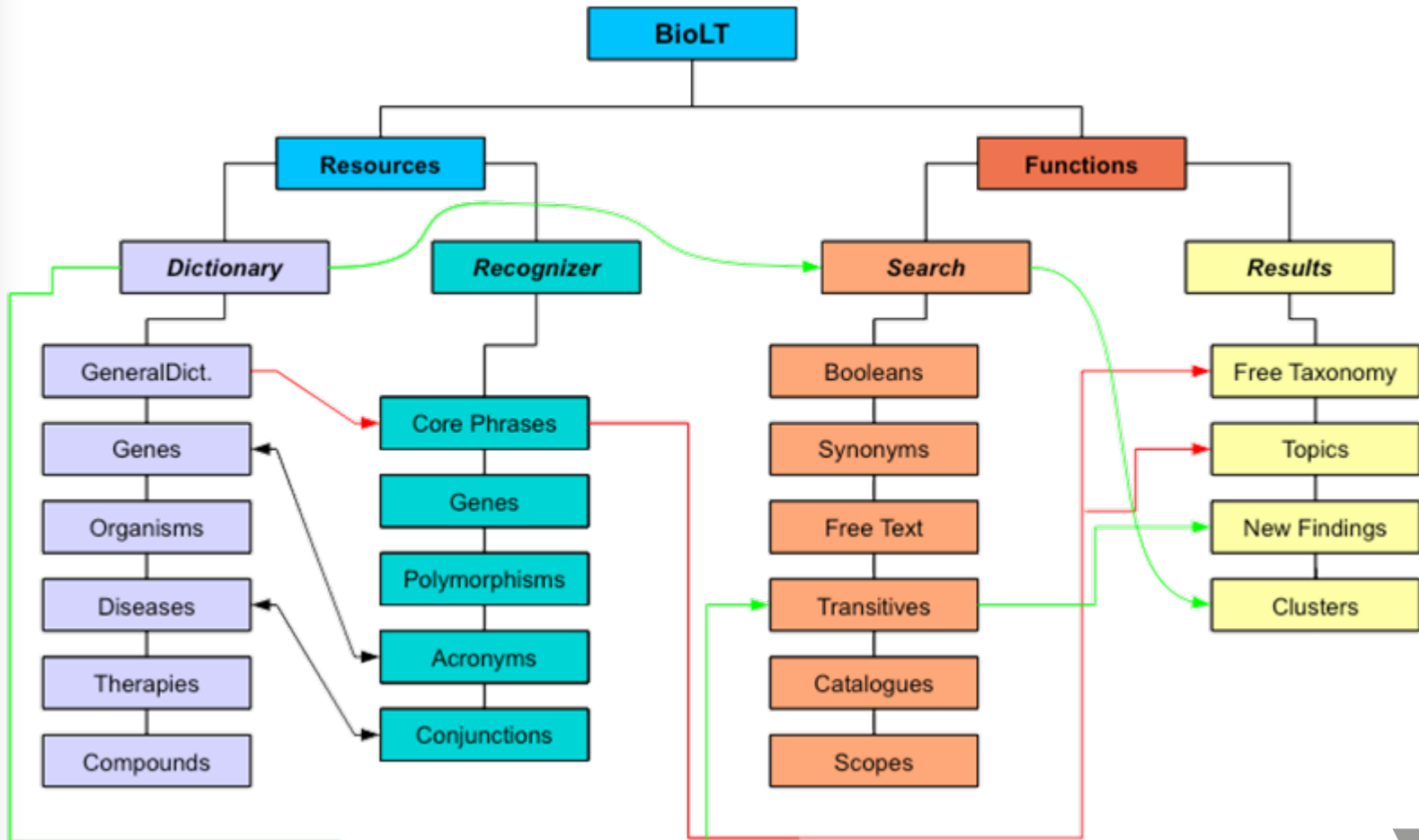


The NCI *Cancer Gene Index* is a database of associations between **genes** and **diseases** and genes and **drug compounds** derived from the biomedical literature as a single source to help cancer researchers to accelerate the search for novel cancer cures.

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

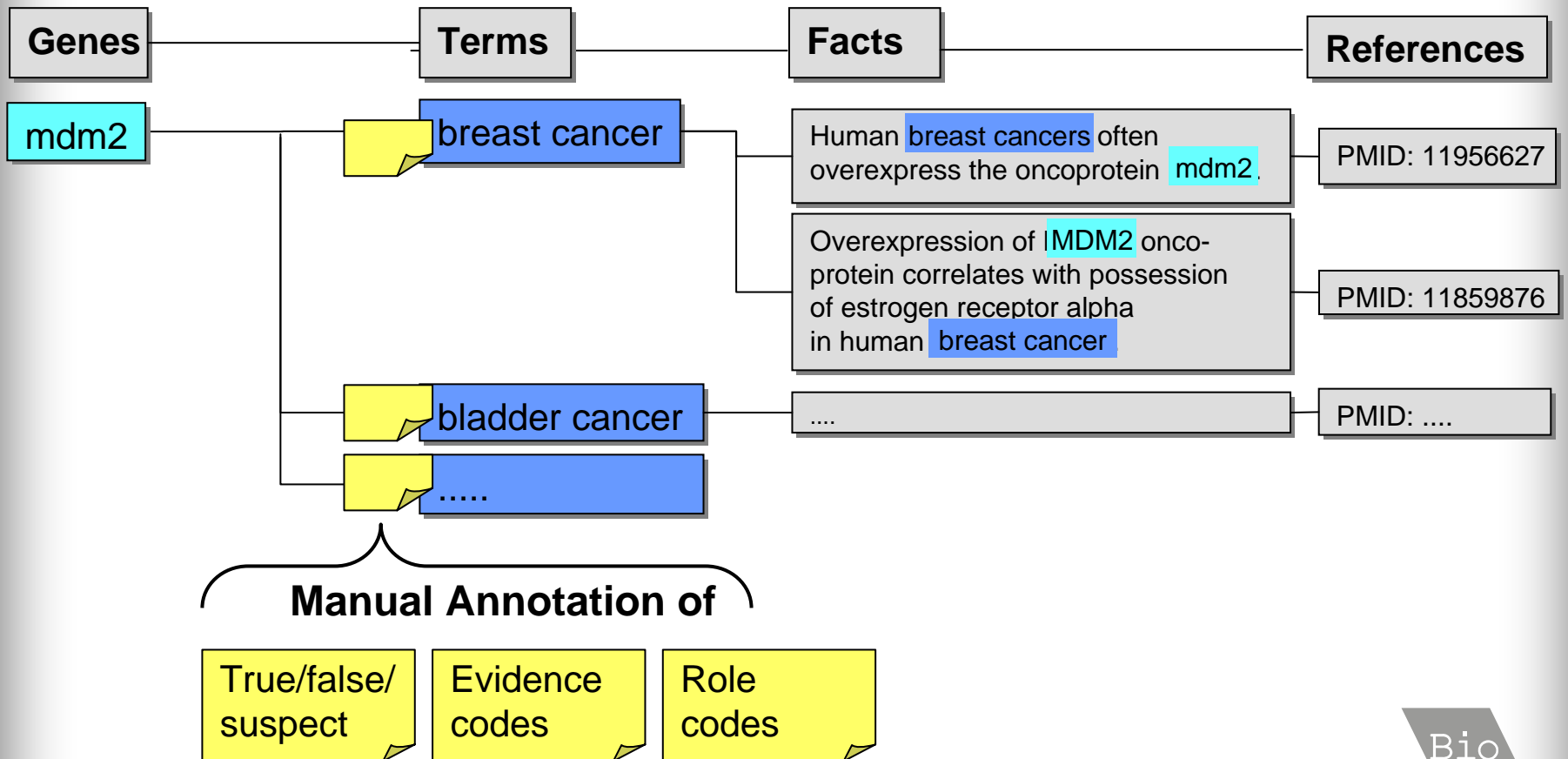
Generating the NCI Cancer Gene Index

1st Step: BioLT Textmining Engine

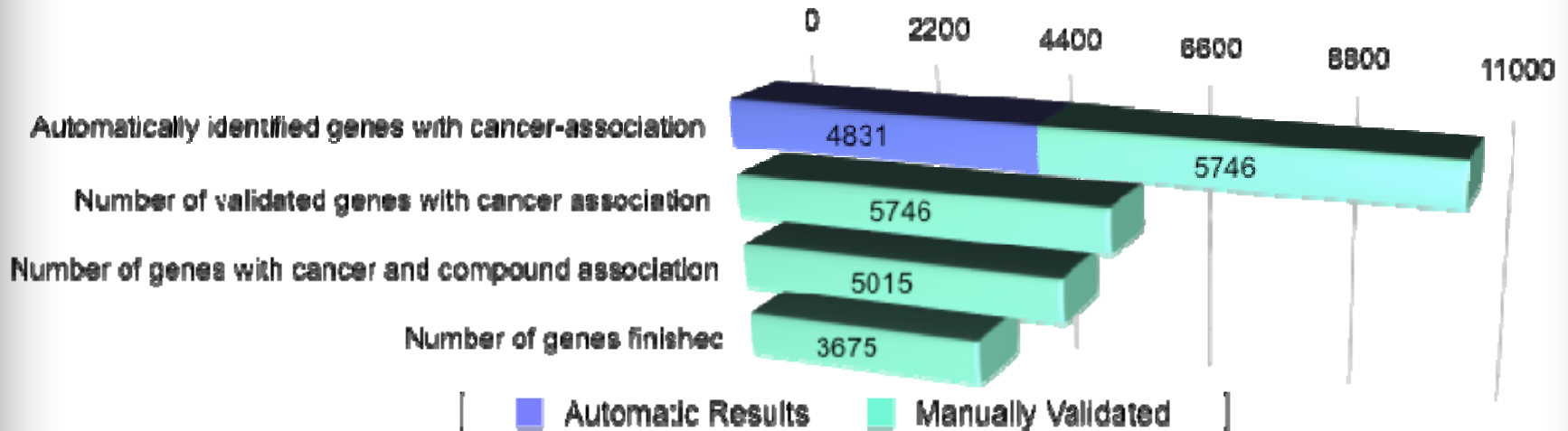


- Selection and configuration of the engine components allows balancing precision and recall
- To generate the NCI Cancer Gene Index, recall was optimized

2nd Step: Manual Curation



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

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

C3868 Gastrointestinal_Stromal_Tumor
 is to
 is a
 C27716 Extragastrintestinal_Gastrointestinal_Stromal_Tumor
 is synonymous to
GASTROINTESTINAL STROMAL TUMOR
 has associated disease
 EV-EXP-IMP Inferred from mutant phenotype.
 EV-EXP-IEP Inferred from expression pattern.
 EV-AS-TAS Traceable author statement.
 it phenotype. sion pattern.
 notype.
 PDGFRA
 has associated disease
 not_assigned not assigned
 EV-EXP-IEP Inferred from ex
 EV-EXP-IMP Inferred from m

Gene-Disease relation 'PDGFRA is associated with disease GASTROINTESTINAL STROMAL TU...

Report Graph Annotations

View: Validated Relation Evidence View

Print... Export to PDF...

Sentence
607546681

Sentence: Gastrointestinal stromal tumors (GISTs) may be defined as intraabdominal nonepithelial (mesenchymal) tumors that express the KIT protein or have an activating mutation in a class III receptor tyrosine kinase gene (KIT or PDGFRA).

PubMed: [15188667](#)

Gene Symbol: PDGFRA

Matched Gene Synonym: pdgfra

Negation: no

Cell Line: no

Evidence Codes: [EV-AS-TAS Traceable author statement.](#)

Role Codes: Gene_Associated_With_Disease
Gene_Product_is_Biomarker_of

Role Details: Gene_Product_Affects_Disease_Process

Organism: Human

Comment: 0

Annotation Status: finished

Help

BioXM – Querying the Oncology Base

“Find genes experimentally associated with specific cancer types”

'Gene-Disease relations based on experimental evidence'

Report Results Annotations

View: Gene-disease relation evidence-frequency

Gene pkd* Search

Gene	Related Disease	Evidenc...	NCI Thesaurus	
			Concept	Inferred Gene...
PKD1	POLYCYSTIC KIDNEY-DISEASE	79	<input type="checkbox"/> C34750 Cystic Kidney Disease	16
PKD1	AUTOSOMAL DOMINANT POLYCYSTIC KIDNEY DISEASE	62	<input type="checkbox"/> C34750 Cystic Kidney Disease	16
PKD1	CYSTS	21	<input type="checkbox"/> C2978 Cyst	45
PKD1	RENAL CYSTS	6	<input type="checkbox"/> C3970 Renal Cyst	2
PKD1	CYSTIC KIDNEYS	3	<input type="checkbox"/> C34750 Cystic Kidney Disease	16
PKD1	CYSTIC KIDNEY	2	<input type="checkbox"/> C34750 Cystic Kidney Disease	16
PKD1	APOCRINE CARCINOMAS	2	<input type="checkbox"/> C4169 Apocrine Carcinoma	1
PKD1	CYSTIC DISEASE	1		

Object to find:

- is a Gene-Disease relation which
 - simultaneously
 - is annotated by Evidence which
 - has Evidence Code attribute which
 - is inferred by ontology entry which
 - has uid
 - like Evidence Code Pattern
 - is owned by 'Biomax Oncology Knowledge Base' project

Kidney Disease	16
Carcinoma	264
Cyst	3

- Project Statistics & Analysis
- Gene-Compound relations based experimental evidence
- Gene-Disease relations based on experimental evidence

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

Evidence 'EV-EXP Inferred from experiment' - Ontology Entry View

Report Derived Concepts Annotations

View: **Statistics for Evidence Codes**

Evidence Code	Evidence	Direct evidence.
EV-EXP-IEP	Inferred from expression	27414
EV-EXP-IDA	Inferred from direct assay.	14444
EV-EXP-IMP	Inferred from mutant	11188
EV-EXP-IPI	Inferred from physical	11007
EV-EXP-IGI	Inferred from genetic	983
EV-EXP-IEP-GENE	Gene expression analysis.	667
EV-EXP-IDA-BINDI	Binding of purified	436
EV-EXP-IDA-BINDI	Binding of cellular extracts.	257
EV-EXP-IDA-PURIFI	Assay of purified protein.	172
EV-EXP-IMP-REAC	Reaction blocked in	195
EV-EXP-IMP-SITE-I	Site mutation.	156
EV-EXP-IMP-REAC	Reaction enhanced in	93
EV-EXP-IGI-FUNC	Inferred by functional	73
EV-EXP-IDA-RNA-IRNA	polymerase	31
EV-EXP-IDA-UNPUI	Assay of unpurified	27
EV-EXP-IDA-TRAN	Length of transcript	16
EV-EXP-IDA-BOUN	Boundaries of	12
EV-EXP-IDA-TRAN	Transcription initiation	7

21 object(s)

View items:

Name	Reports
Evidence Code	Ontology entry Uid
Evidence	Name
Direct evidence classifications	Number of query results
Directly classified relations	Number of query results
Inferred evidence classifications	Number of query results
Inferred relations	Number of query results

Step 1: Define Query

Object to find:

- Is a relation which
 - is annotated by Evidence which
 - has Evidence Code attribute which
 - is inferred by ontology entry which
 - is in Set of Evidence Codes

Delete

is annotated by ...

Annotation form:

Evidence

Apply Discard

Variable definitions:

Name	Type	Default value
Set of Evidence Codes	Object Set	

Add... Edit... Delete

BioXM – Table-driven Knowledge Processing

Step 1: Define import operations

Stored sequences:

Available import operations:

- Create/Lookup
 - Lookup or Create element
 - Lookup or Create context
 - Create relation
 - Create annotation
- Add object to context
- Add object to folder
- Add relation to context
- Assign alias to object
- Assign annotation
- Assign description
- Assign experimental data
- Assign sequence
- Change annotation
- Create BioRS mapping
- Delete object
- Delete object alias
- Detach annotation

Import operation sequence:

<input checked="" type="radio"/> Lookup or Create element	\$element1
Name	A
Type	<input checked="" type="checkbox"/> SNP
Create in project	Current
<input checked="" type="radio"/> Create annotation	\$annotation1
Object	\$element1
Name	Empty
Form	<input checked="" type="checkbox"/> SNP Prop...
Overwrite	Yes
Create in project	F
Attributes	
Validation status (dbSNP)	E
Alleles	D
Mutation	C
SNP Quality	No Value
SNP position	No Value

Lookup or Create element

Result object local name:

Description:

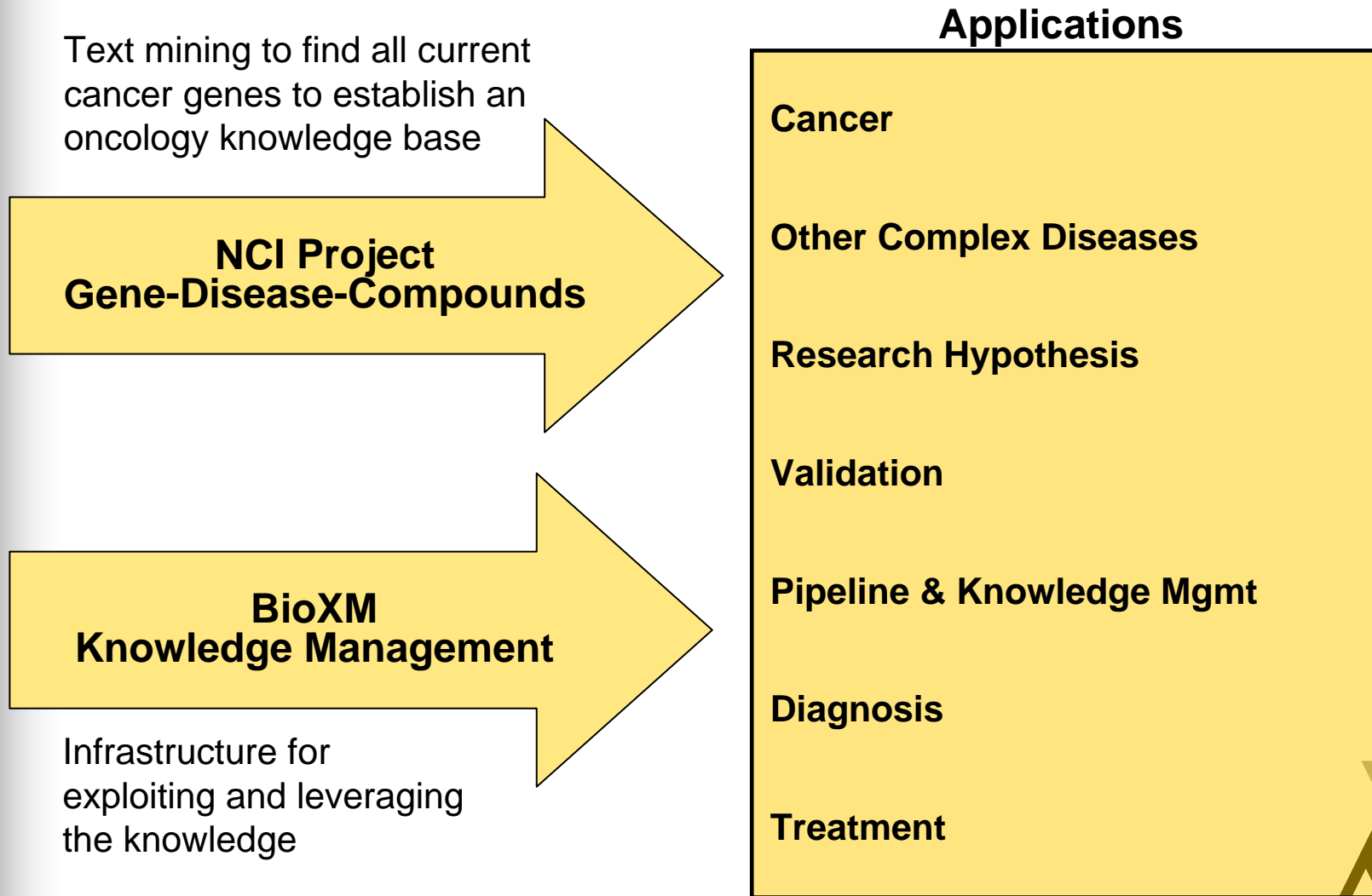
This import operation searches for element(s) of the specified type and name. In case an element is not found, it is created.

The result of this operation is the found or created element(s) reference which can be used as an argument for further operations.

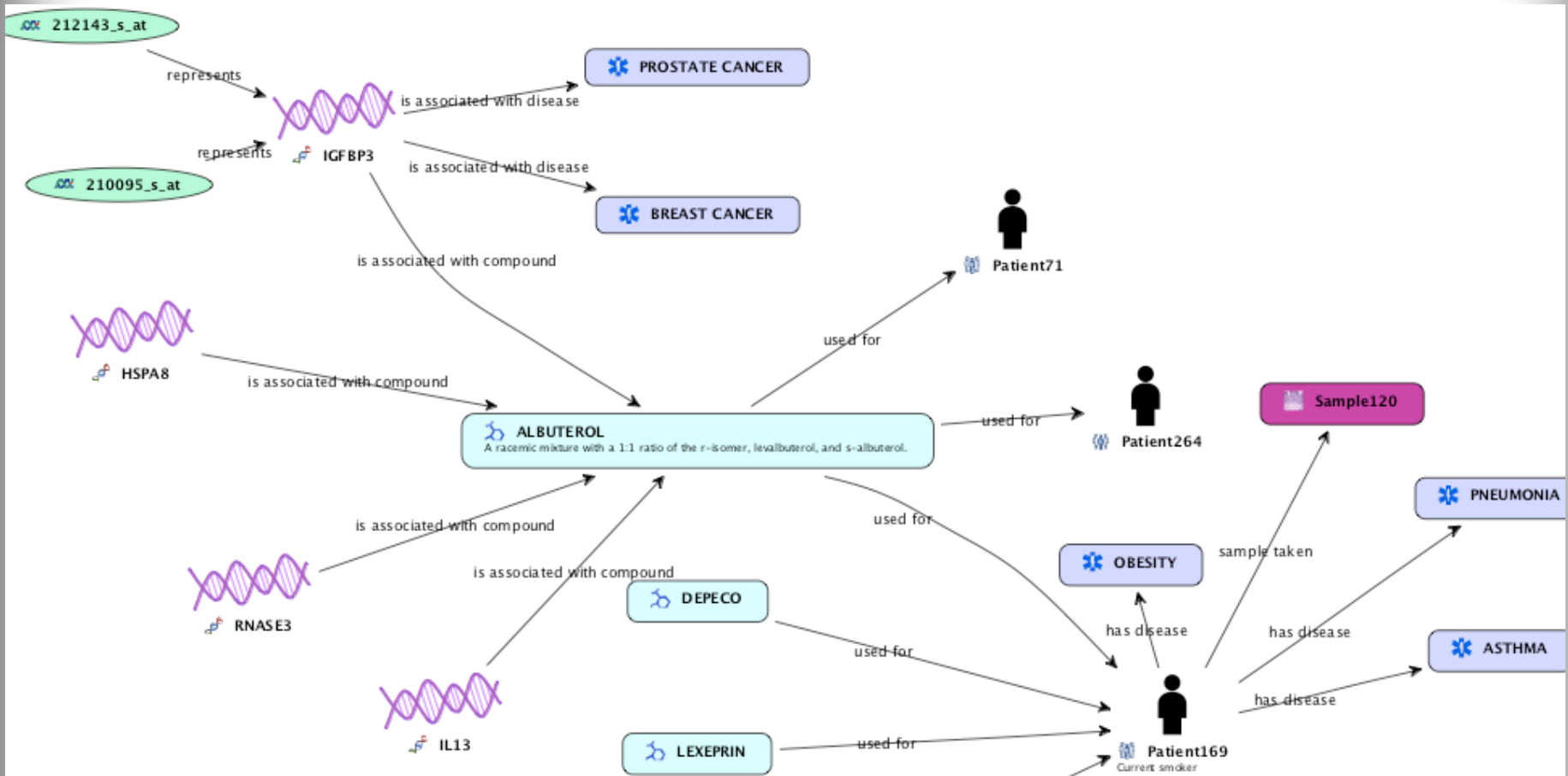
Data preview:

	A	B	C	D	E	F
1	SCN5A_p278ca	SNP	p278ca(1p114ca)			SNP Analysis
2	rs11601907	SNP	Y662Y	C/T	Y	SNP Analysis
3	rs1137617	SNP	Y652Y		N	SNP Analysis
4	SCN5A_V728I	SNP	V728I			SNP Analysis
5	rs4680	SNP	V108M	A/G	Y	SNP Analysis

Conclusion



Thank you!



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