Knowledge management in a Systems Biology approach to translational medicine

Dieter Maier[1], Miguel Hernandez[2], Jordi Villa i Freixa[2] and Sascha Losko[1]

1 Biomax Informatics AG ● Lochhamer Str. 9 ● 82152 Martinsried ● Germany ● bioxm@biomax.com
2 Universitat Pompeu Fabra ●Placa de la Merc, 10-12 ● 08002 Barcelona ● Spain

Introduction

The **EU BioBridge project** approaches complex phenotypes on a molecular level which requires computational modelling and simulation based on specific environments.

- 7 partners from academia, clinic and industry
- Mechanisms of chronic obstructive pulmonary disease (COPD), cardiovascular diseases and diabetes
- Chronic systemic inflammation as common disease mechanism?
- Biomarker guided diagnosis and therapy suggestions
- Simulation based phenotype analysis, diagnosis and therapy
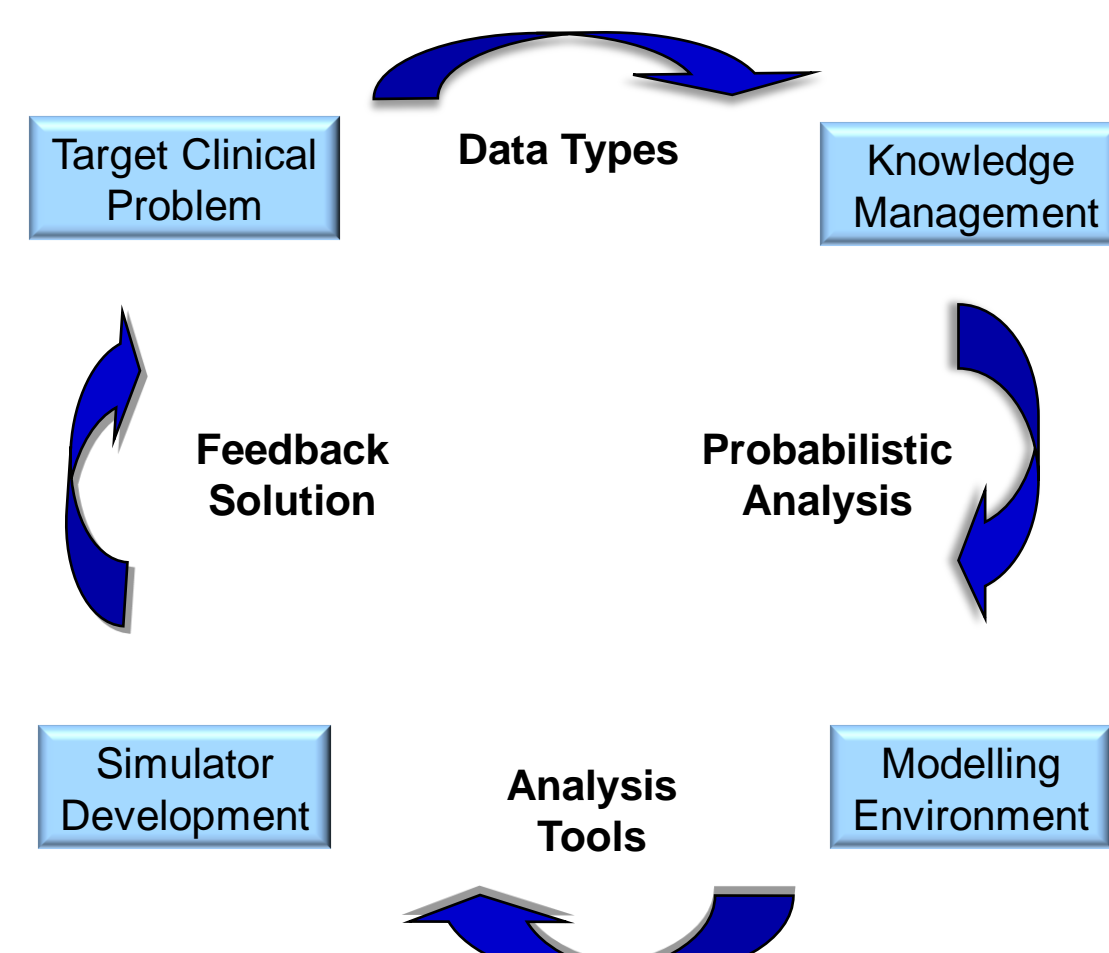
**Why Knowledge management**?

- Judge biological significance of statistical data analysis
- Adapt general models to specific environments
- Connecting different levels of knowledge (molecular, physiologic, clinical) is a prerequisite for a translational approach to research

What are the challenges?

- Joining different levels of information in a system requires objects highly connected in different scientific fields
- Mapping entities e.g. genes from different sources (Genbank ID, KEGG ID, Literature name)
- Mapping descriptions e.g. cancer - blastoma by using controlled, structured vocabularies
- Navigating the network resulting from data integration and mapping, providing suitable sub-networks to answer questions
- Reporting exactly all significant parts of the sub-network answering the question
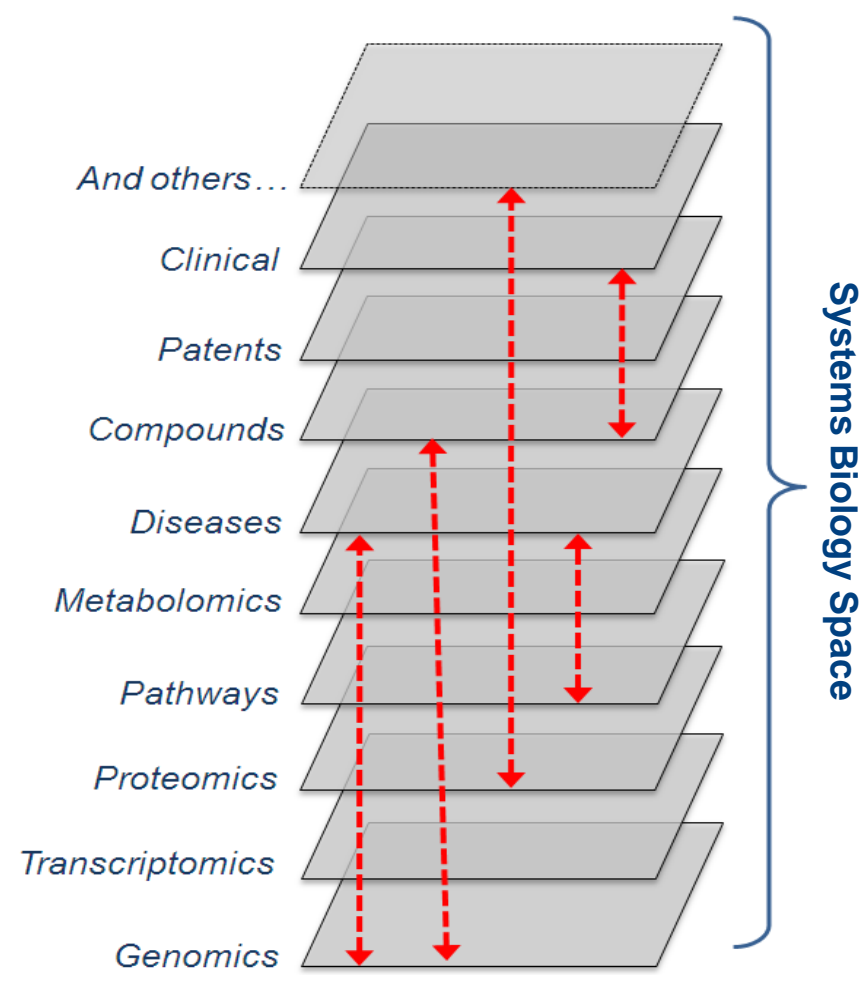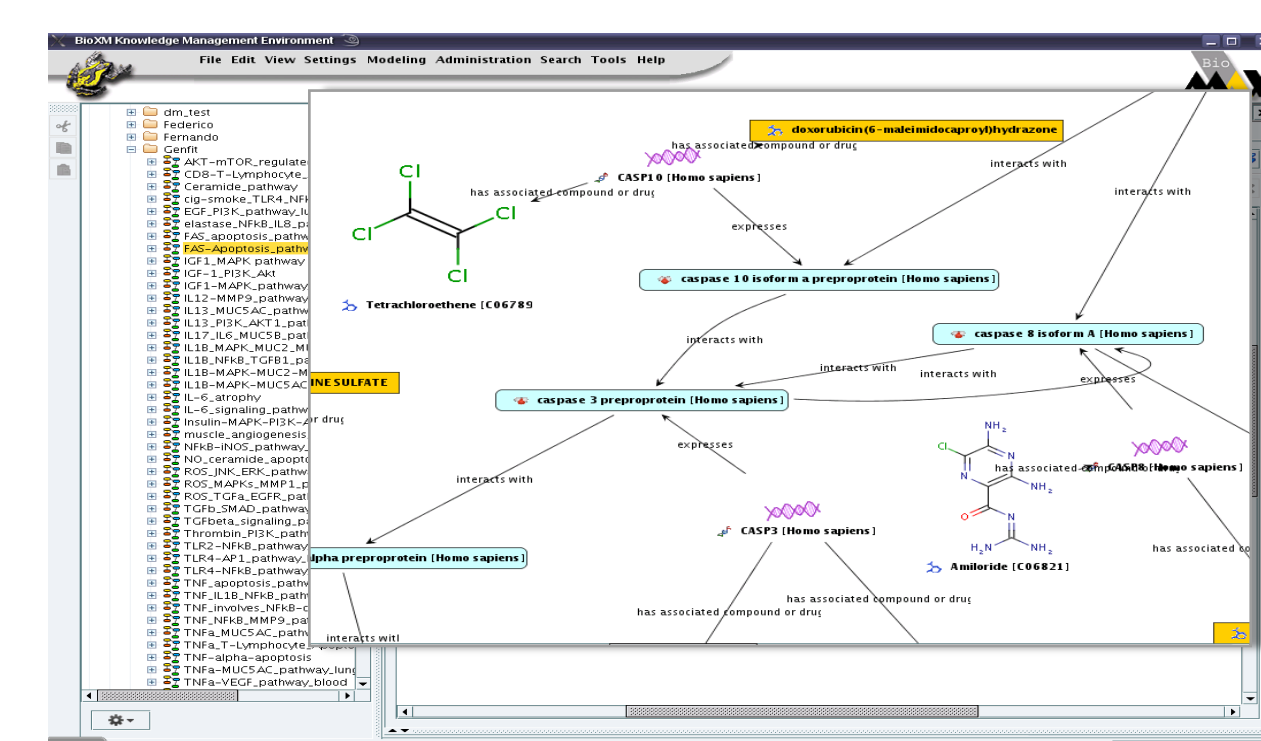
## Object oriented semantic mapping and dynamic reporting
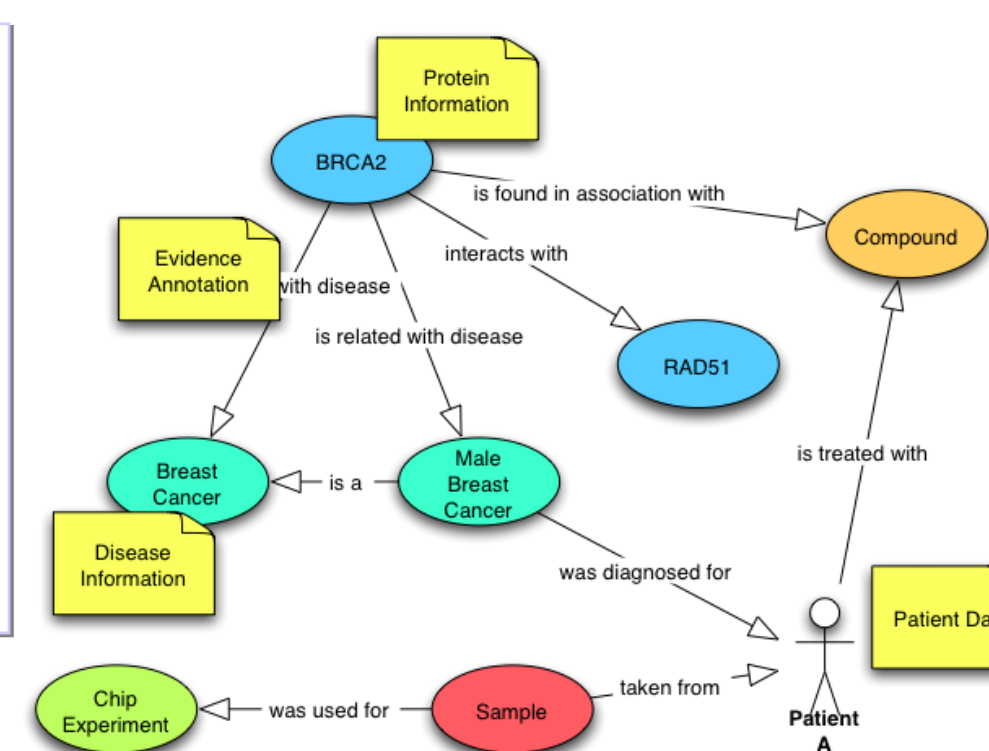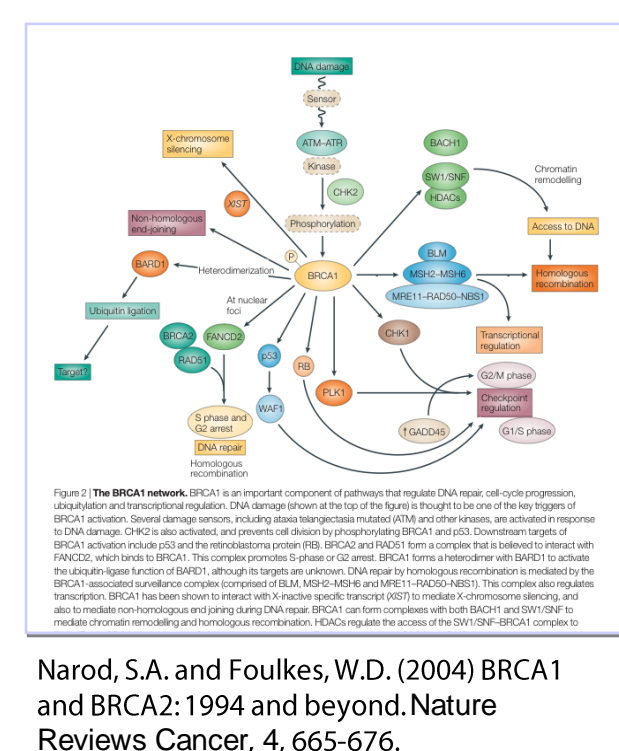
BioXM conceptualises the standard representation of life science knowledge, the scheme of interacting entities and their relations. Based on controlled vocabularies and dynamically defined objects, relations and annotations the data is integrated into a project specific knowledge network.

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

Information can be searched, represented and reported in multiple ways
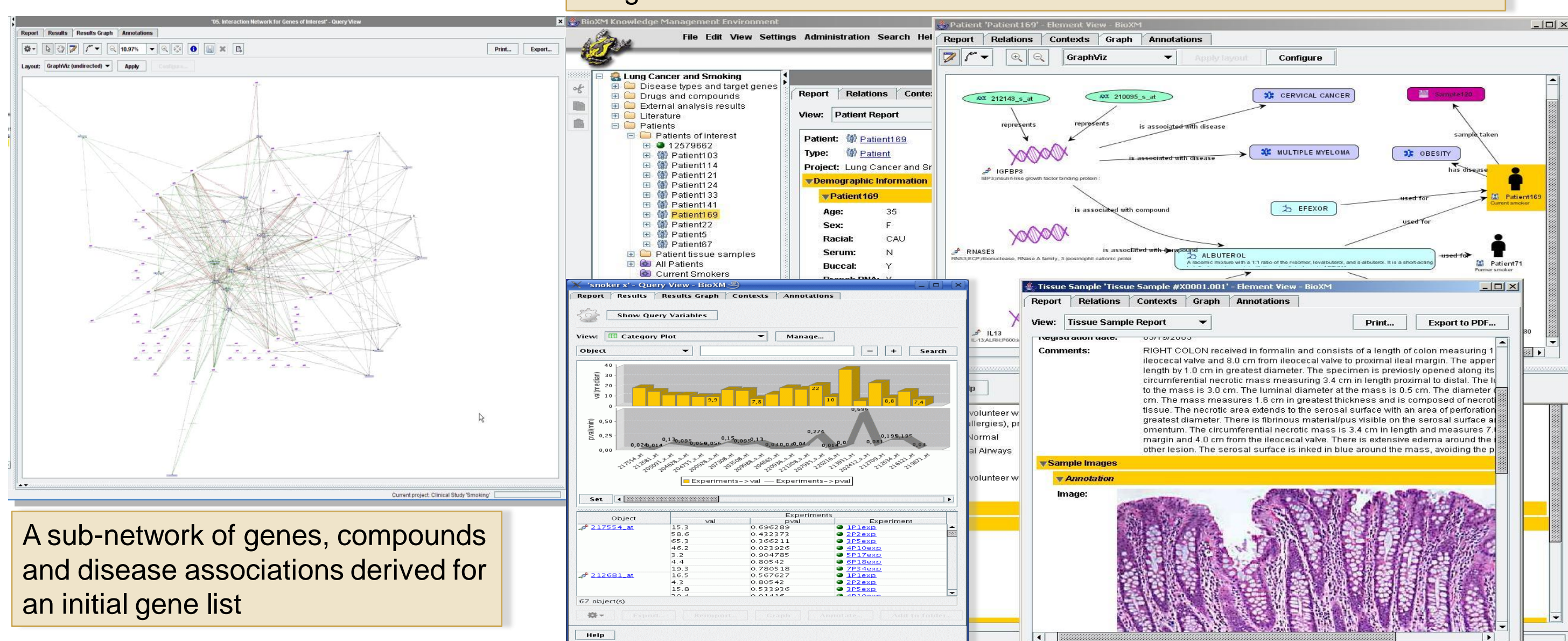
Tabular reports associate objects with dynamically defined information here heatmap, cluster and expression value
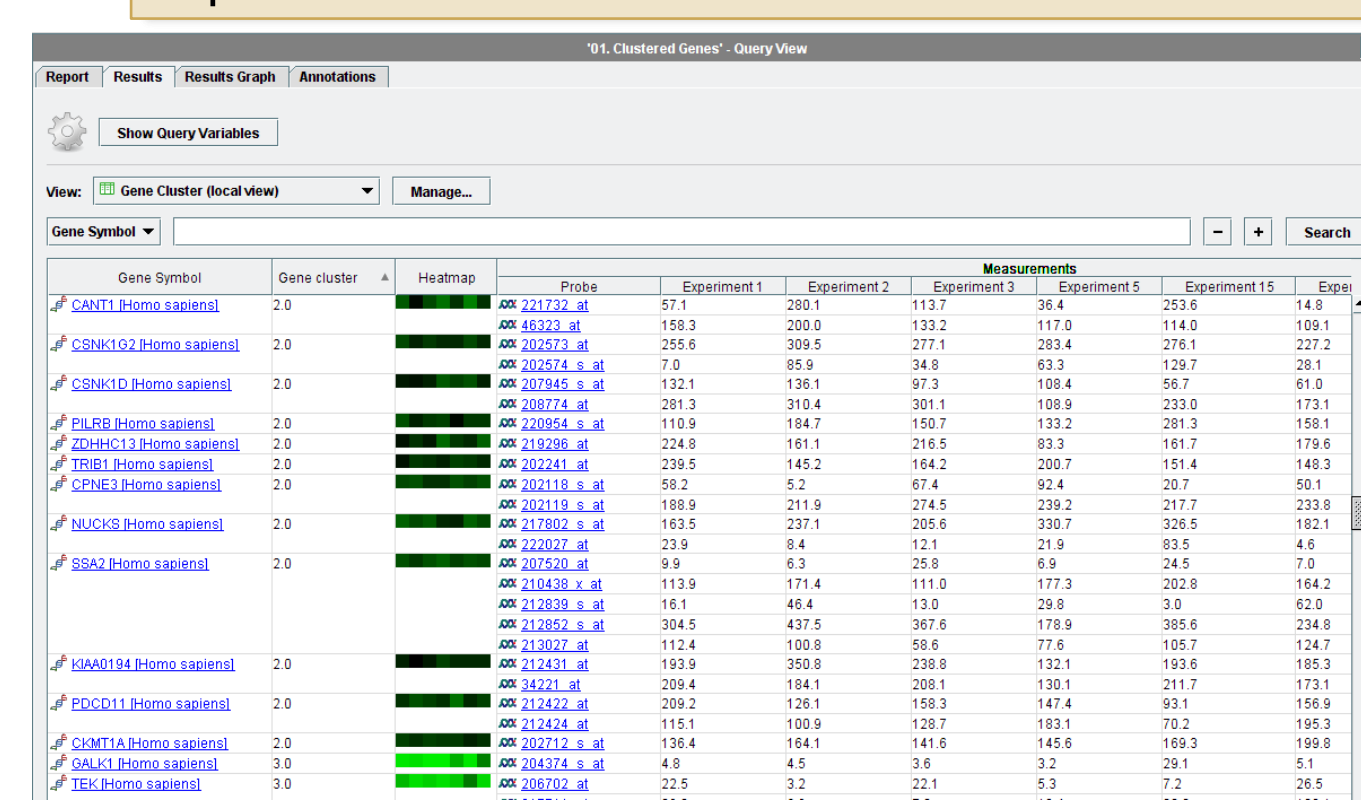
Literature derived tissue specific FAS pathway integrated with gene-compound data from the Comparative Toxicogenomics Database

Multiple object reports, textual, graphical, as histogram, including images

A sub-network of genes, compounds and disease associations derived for an initial gene list

Integrating COPD specific molecular networks, clinical and experimental data

The BioBridge BioXM instance integrates
(directly or by interfacing)

- > 20 public databases
- > 10 ontologies
- 80 793 genes (30 246 human, 27 237 mouse, 23 310 rat)
- 1 307 pathways
- 78 528 compounds
- 1 525 474 protein interactions
- 3 666 313 connections overall in the knowledge network
- The entire Gene Expression Omnibus and BioModels databases
- BioBridge internal clinical and experimental data (expression, metabolomics and proteomics)
- BioBridge literature-mining derived molecular networks for COPD, cardiac disease, chronic systemic inflammation, diabetes and lung and muscle specific signalling sub-networks.
- Literature derived exercise and COPD specific kinetic and metabolic data
- Mathematical models and probabilistic networks generated within the BioBridge project
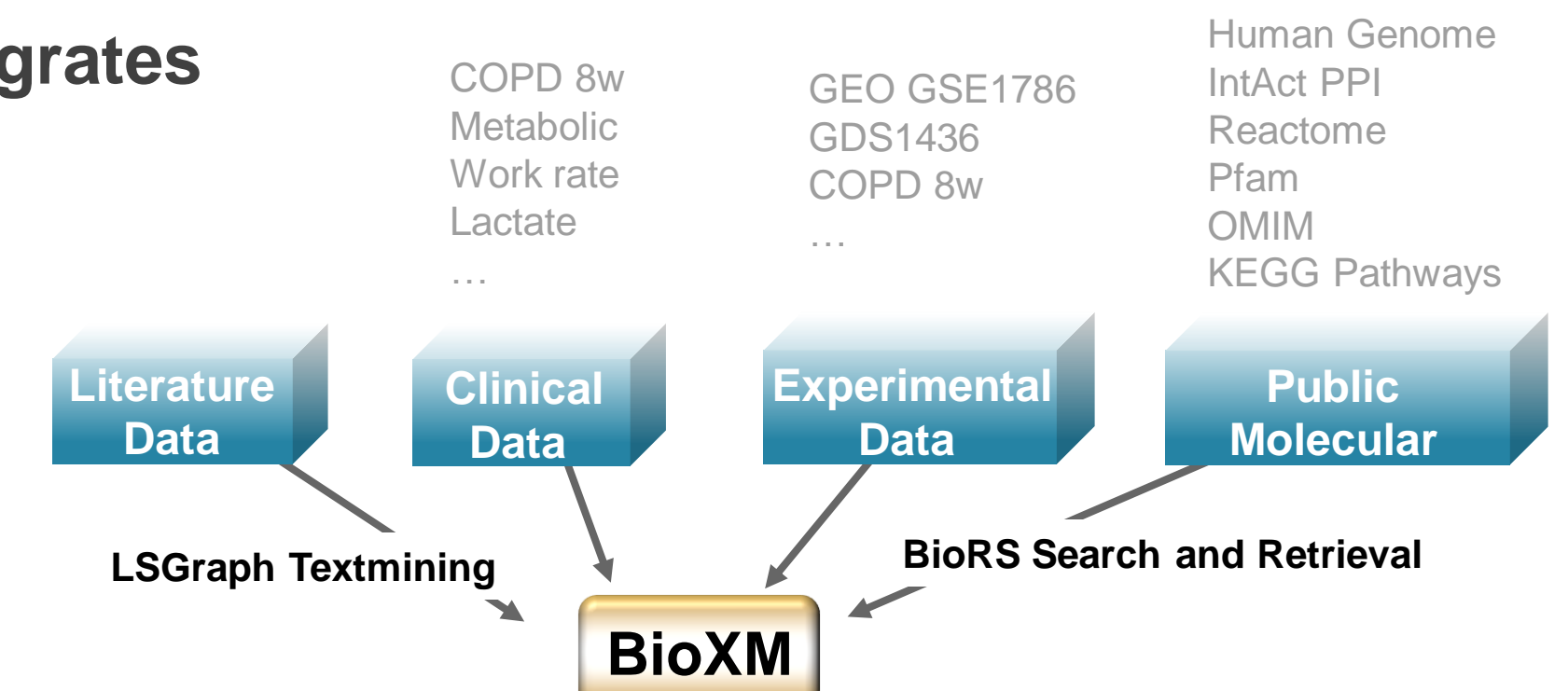
Partial view of the BioBridge BioXM data model
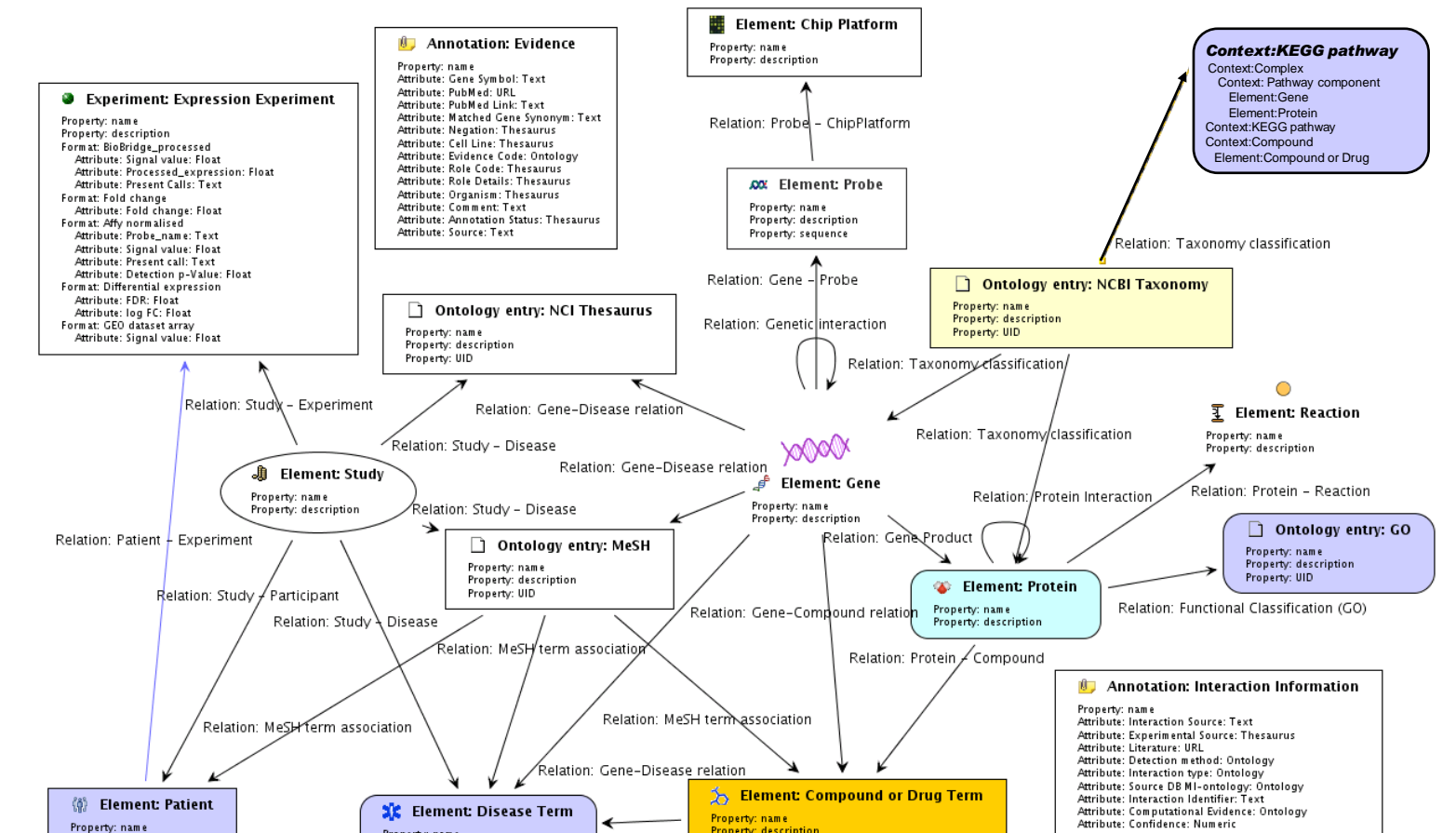
Extracting sub-networks of integrated information for modelling and simulation

Context specific sub-networks such as the muscle specific remodelling pathways are integrated with the corresponding experimental data. Data analysis such as principal component analysis (PCA) for pathways or seed based network inference find and extend molecular networks affected by condition and treatment

Pathway PCA analysis Close-up visualises impact of individual gene to PCA

Context specific Sub-Networks
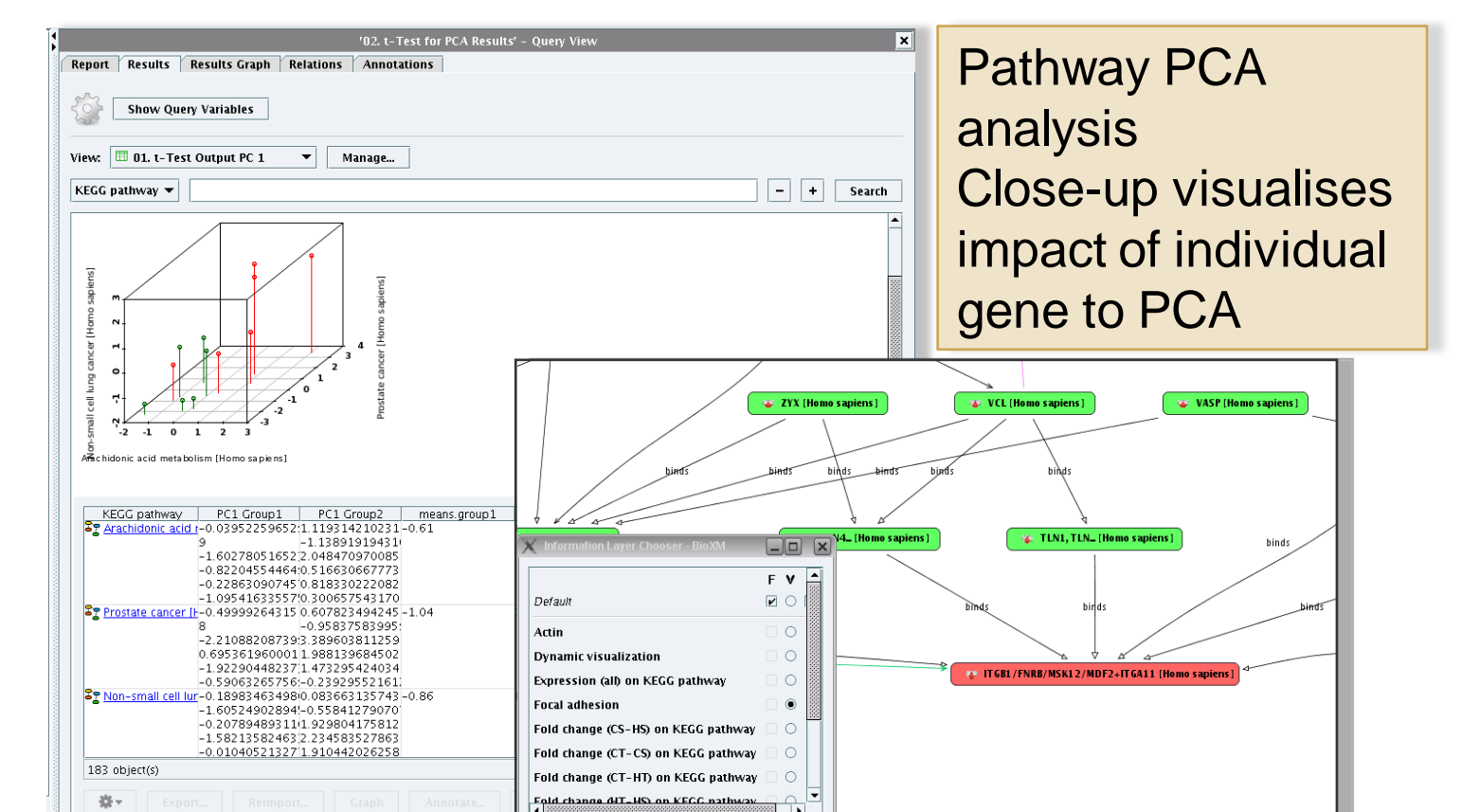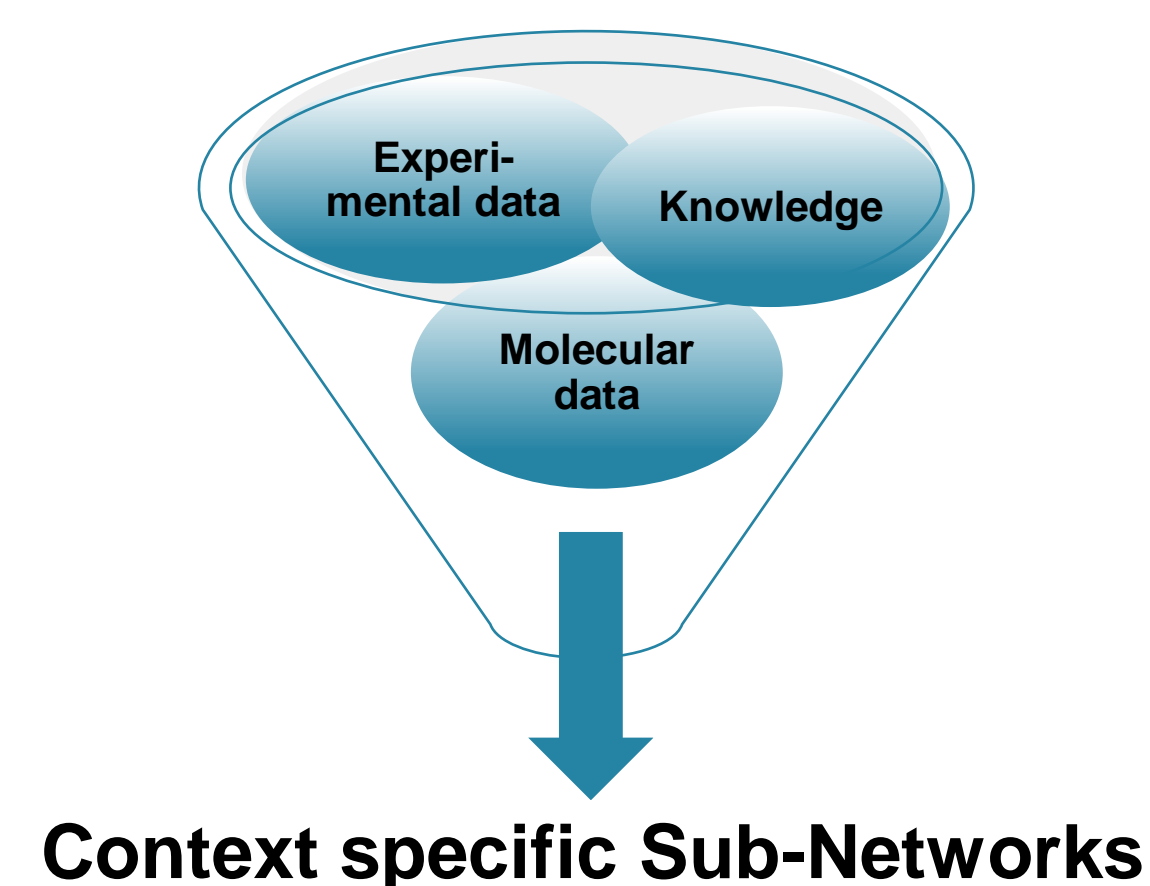
Conclusion

The BioXM knowledge management system, integrated into the BioBridge portal, provides the multi-level knowledge, data integration and retrieval required to model complex phenotypes such as COPD and to iteratively improve simulation results and experimental approaches.
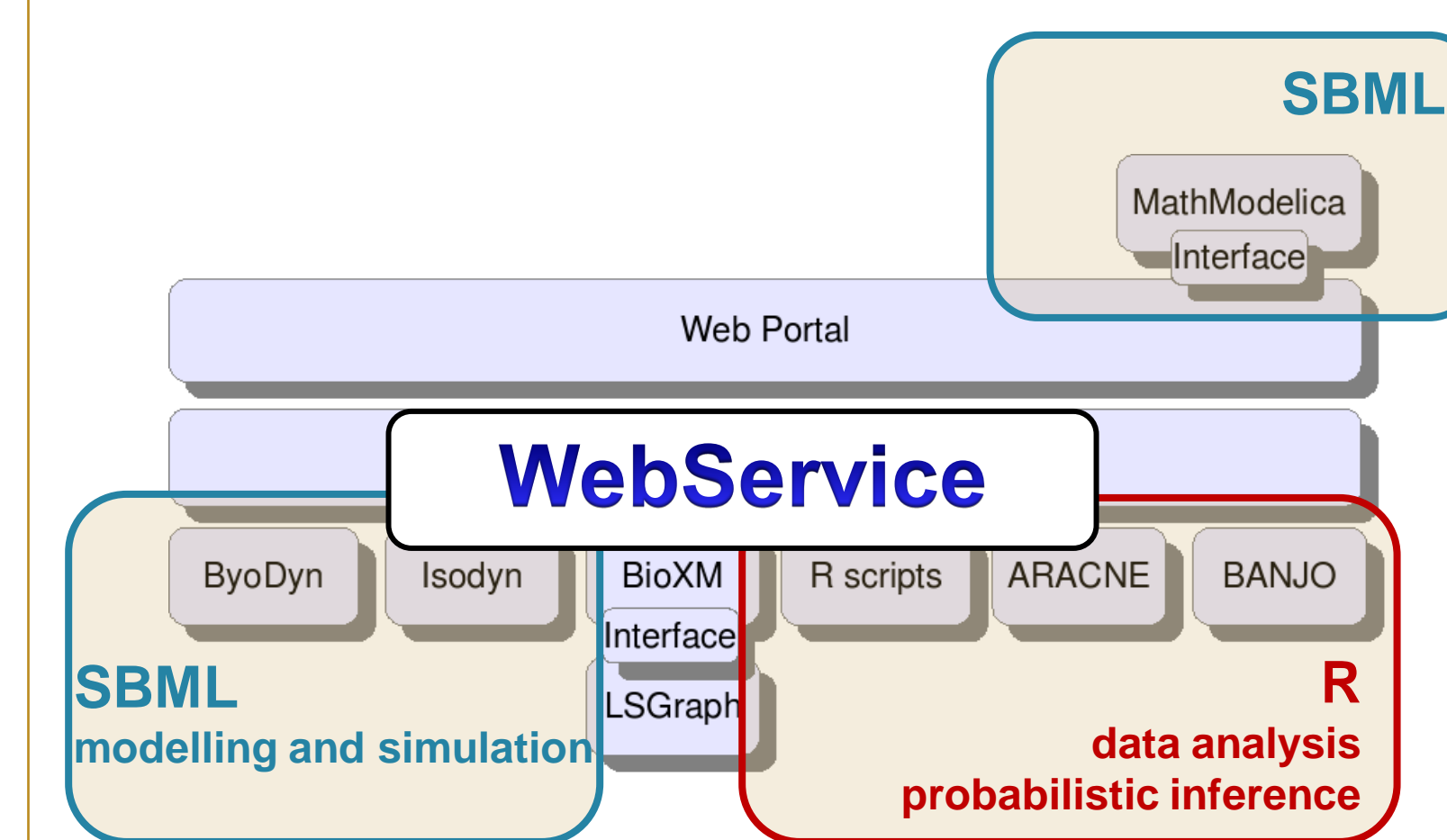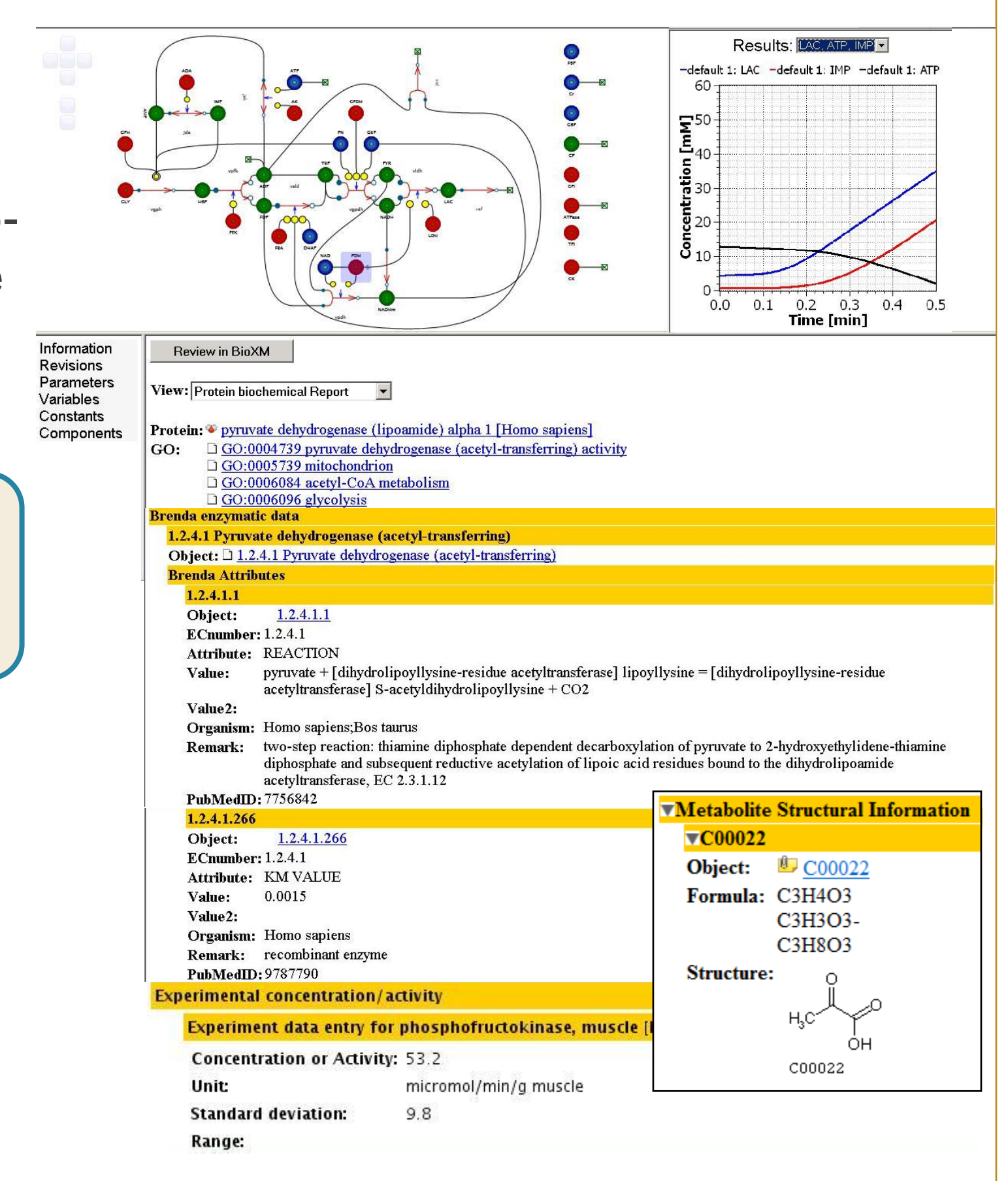
References

Selivanov VA, et al., (2008). The changes in the energy metabolism of human muscle induced by training. Journal of Theoretical Biology 252, 402-410

Sameith, K. et al., (2008). Functional Modules integrating essential cellular functions are predictive of the response of leukaemia cells to DNA damage. Bioinformatics 24: 2602-2607

Losko, S. et al. (2006) Knowledge Networks of Biological and Medical Data: An Exhaustive and Flexible Solution to Model Life Science Domains. Lecture Notes in Computer Science 4075: 232-239

Acknowledgement

www.biobridge.eu   www.biomax.com

BIOBRIDGE

biomax
biomax informatics ag