Deliverable No. D16.4

Training Course on the Literature Based Discovery platform

Grant Agreement No.:	270089
Deliverable No.:	D16.4
Deliverable Name:	Training course on the Literature Based Discovery Platform
Contractual Submission Date:	31/01/2014
Actual Submission Date:	31/01/2014

Dissemination Level		
PU	Public	X
PP	Restricted to other programme participants (including the Commission Services)	
RE	Restricted to a group specified by the consortium (including the Commission Services)	
CO	Confidential, only for members of the consortium (including the Commission Services)	

COVER AND CONTROL PAGE OF DOCUMENT	
Project Acronym:	***p-medicine***
Project Full Name:	From data sharing and integration via VPH models to personalized medicine
Deliverable No.:	D 16.4
Document name:	Training course on the Literature Based Discovery platform
Nature (R, P, D, O)[1]	D
Dissemination Level (PU, PP, RE, CO)[2]	PU
Version:	1.1
Actual Submission Date:	04/02/2014
Editor: Institution: E-Mail:	Anuj Sharma Biovista a.sharma@biovista.com

ABSTRACT:

This document provides training material for the usage of an LBD tool, the Correlation Viewer, and the application of LBD processes to scenarios of interest to Clinicians and Bioinformaticians.

KEYWORD LIST: Literature Based Discovery, Correlation Viewer, Training material

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 270089.

[1] **R**=Report, **P**=Prototype, **D**=Demonstrator, **O**=Other

[2] **PU**=Public, **PP**=Restricted to other programme participants (including the Commission Services), **RE**=Restricted to a group specified by the consortium (including the Commission Services), **CO**=Confidential, only for members of the consortium (including the Commission Services)

MODIFICATION CONTROL			
Version	Date	Status	Author
0.1	18/01/2014	Draft	Miltiadis Koutsokeras
0.2	27/01/2014	Draft	Anuj Sharma
1.0	30/01/2014	Final	Anuj Sharma
1.1	04/02/2014	Final	Anuj Sharma

List of contributors

- Militiadis Koutsokeras, Biovista
- Anuj Sharma, Biovista
- Danny Burke, ecancer

Contents

1 Executive Summary

This document provides training material for the usage of an LBD tool and the application of LBD processes to scenarios of interest to Clinicians and Bioinformaticians.

The document contains detailed descriptions on the usage of the Correlation Viewer, which is an LBD tool developed by Biovista for the p-medicine project. Further, scenarios of clinical interest are developed and analyzed with an emphasis on the application of LBD for drawing conclusions.

2 Introduction

The purpose of this document is to present a tool and processes, available to Bioinformaticians and Clinicians, for the application of Literature Based Discovery (LBD) to their needs. A detailed description of a LBD tool, Correlation Explorer, developed by Biovista for the p-medicine project is provided in the document. Further, clinical scenarios for using the tool are developed and discussed as a way to elucidate the process for the tools application.

Objectives

The primary objective of this document is to supplement the overall educational goals of the p-medicine project. Through a description of the tool and processes for application of LBD the document aims to contribute to the following objectives:

- Broaden the pool of competent LBD users
- Support the activities and tools of p-medicine

Key Audiences

The audience segments being targeted by the p-medicine project have been defined in D2.2. These target groups range from Healthcare providers to patients. The material discussed in this document is targeted towards p-Medicines internal audiences.

- Clinicians and Clinical research organizations
- Bioinformaticians
- IT professionals

The document aims to provide a resource for improving competence within the project partners for better usage of LBD tools as well as documenting the functionality of the Correlation Viewer.

In order to provide a content-rich training course, for assisting in application of LBD processes to Clinical scenarios, a video tutorial detailing the Correlation Explorer has also been developed. The video tutorial is currently accessible through ecancer's ftp site (http://share.ecancer.org/Pmed/Biovista/), the video will be made available in ecancer's personalised medicine area the ecancer.org platform which will host all of the educational tools developed as part of the p-medicine project and is currently in development.

3 Correlation Viewer

Application usage

The main goal of this application is to be as user friendly as any well-known search engine application, like Google, Yahoo and Bing. The main entry screen of the currently deployed version is shown in the following figure:

Figure 1: Correlation viewer entry screen

The user interface is simple and self-explanatory: a single text field on the center is the main user input entry, and as it is shown it expects the user to type a drug name by default. The list field on the left can be used to determine the type of queries the text field accepts, which is Drugs, Diseases or Pathways:

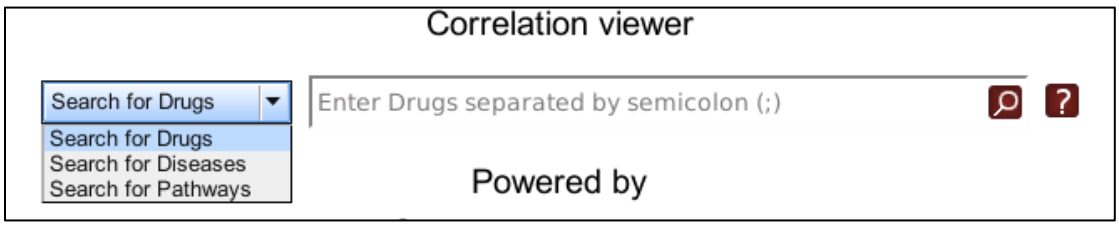


Figure 2: Supported query types

When the user starts typing his query the application switches to the results layout:

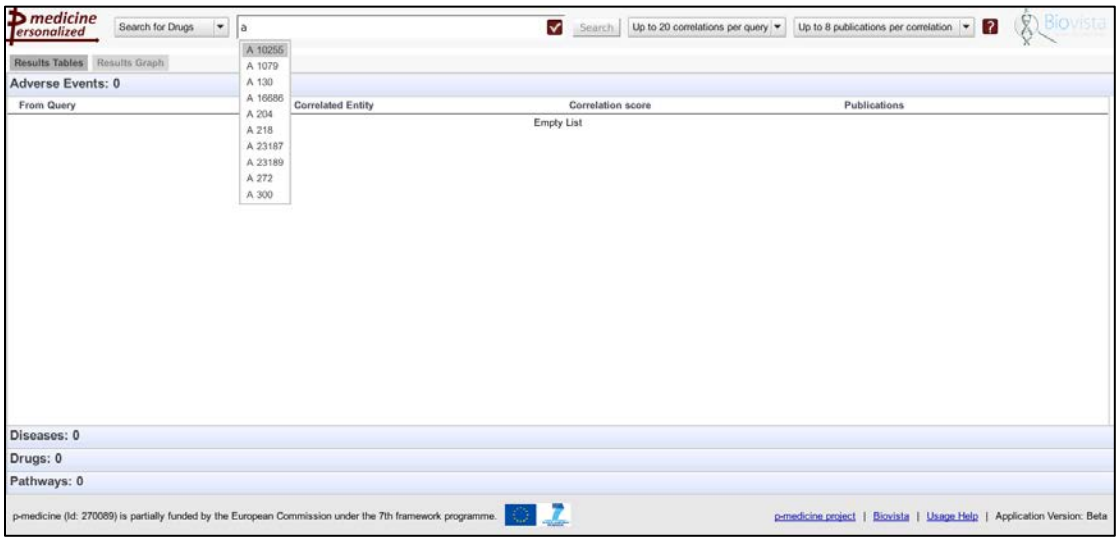


Figure 3: Correlation viewer results screen

In order to minimize the requests of mistyped terms or terms nonexistent in the database, the field responds to user typing with a list of suggested queries:

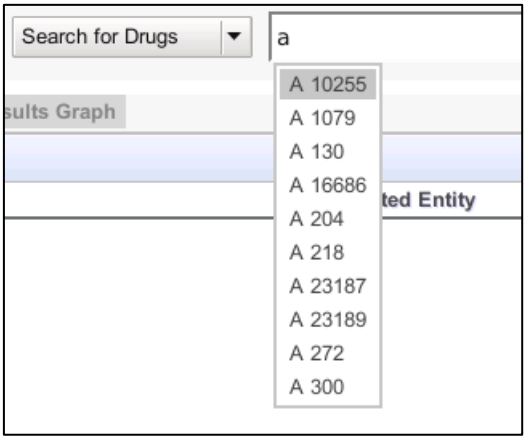


Figure 4: Queries suggestion

The user then can either select one of the suggestions if it matches his query, or continue typing to get a new suggestion list. It's not possible to submit an arbitrary query, not proposed by a suggestion list. When a suggestion is chosen from the list, it is appended on the search text field and the semicolon ';' character is inserted at the end. This character is used to separate multiple query terms from each other. After the append action the user can continue to type the next query term in order to get a new suggestion list:

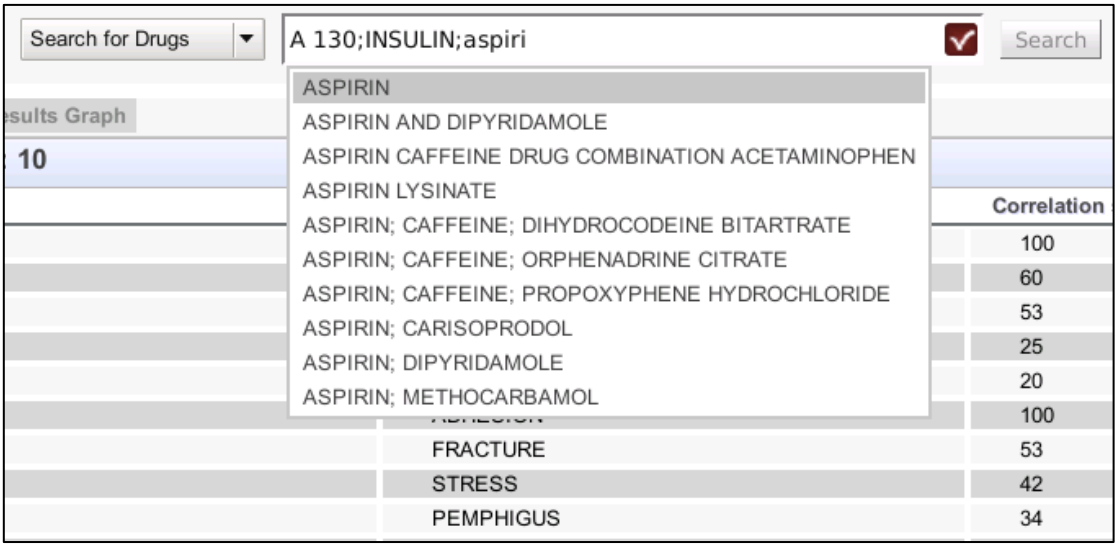


Figure 5: Multiple queries

This procedure can be repeated to create a multi-query of preference. Take note that larger query lists will take longer for the service to process and will result in a result set that can be overwhelming to the user. Now let's perform a search example to get accustomed with the results screen. We will use the drug "Methotrexate", which returns the following response:

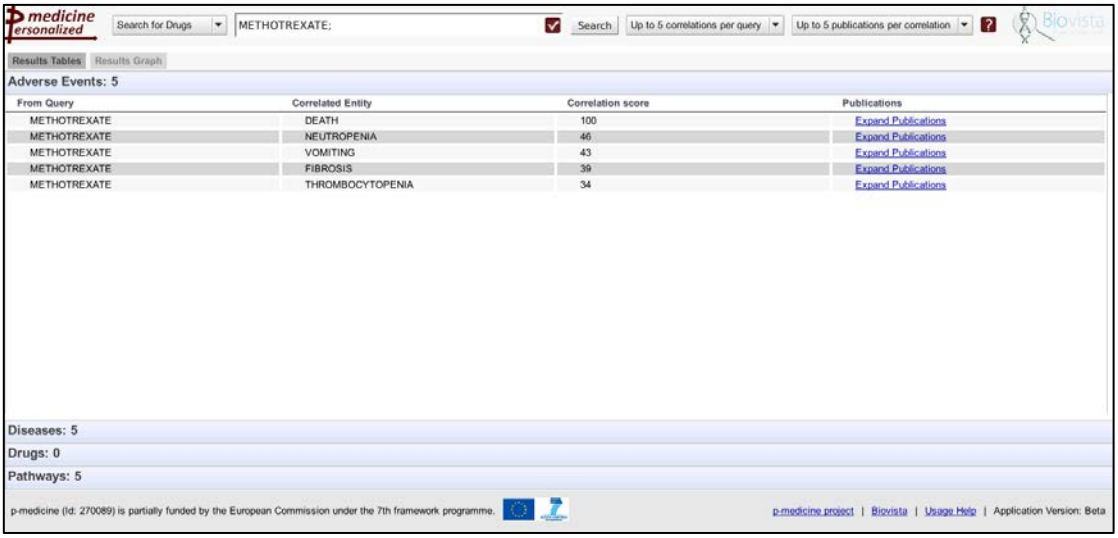


Figure 6: Results example

From the first look we can see that our criteria returned five adverse events, no drugs, five diseases and five pathways. The output is split into four different panels, one for each type of result set. To review some other type of result set we click on the panel of the appropriate type:

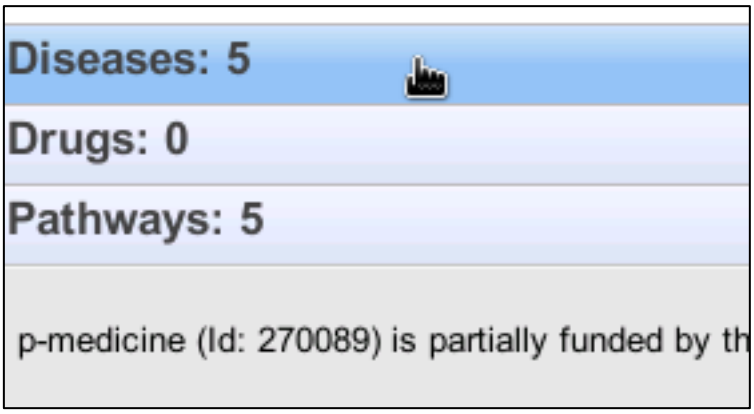
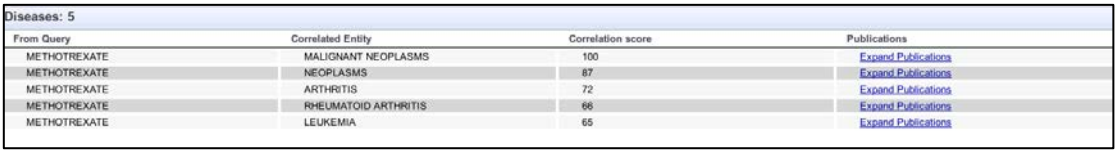


Figure 7: Click on the appropriate panel

We can view the result table of preference:


Diseases: 5

From Query	Correlated Entity	Correlation score	Publications
METHOTREXATE	MALIGNANT NEOPLASMS	100	Expand Publications
METHOTREXATE	NEOPLASMS	87	Expand Publications
METHOTREXATE	ARTHRITIS	72	Expand Publications
METHOTREXATE	RHEUMATOID ARTHRITIS	66	Expand Publications
METHOTREXATE	LEUKEMIA	65	Expand Publications

Figure 8: Diseases result table shown

Each table row is a correlation result and is comprised of four columns. The first indicates the query term that returned this row. This column informs the user of the specific term that returned the row in a case of multiple queries. The second column indicates the term that is correlated with the query of the first column. The third column is the score number of this correlation, the higher the value, the stronger the correlation is. This is the normalized value, to the domain [0, 100], of the number of co-occurrences these terms have in literature. 100 denotes the strongest correlation in the current result set, while 0 means the weakest. To review the associated literature we need to click on the **"Expand Publications"** link of the last column:

Figure 9: Click to expand publications list

The publication list is expanded just below the table row:

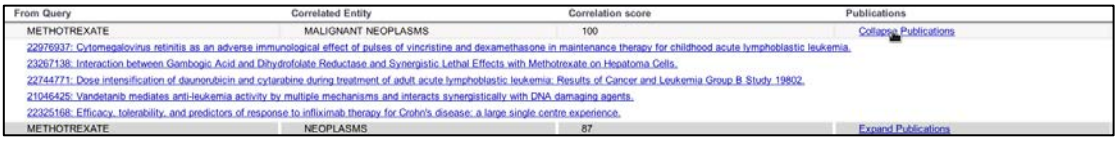


Figure 10: Expanded publications list

To retrieve more results you can change the limits by using the list right of the **"Search"** button:

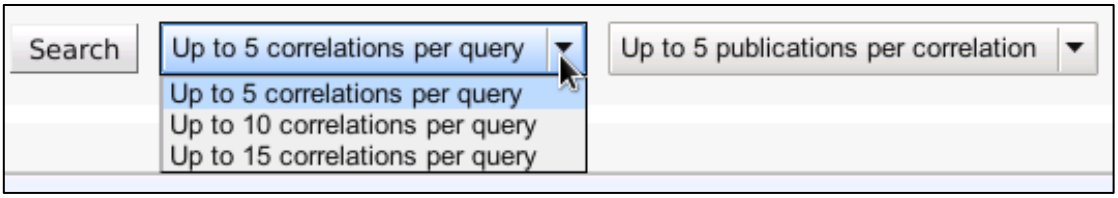


Figure 11: Choose result set limits

Increasing these limits will return larger results sets in the expense of increased retrieval time.

Along with the table panels' view, the application supports a graph representation of the result set. To switch to this view click on the **"Results Graph"** tab near the screen top:

Figure 12: Click tab to switch to graph view

The graph view displays the queries, correlations and publications as colored nodes. The red nodes on the graph are the queries, linked with their corresponding correlations, colored in a per type basis. Publications are linked with each correlation node. At the left there is a graph legend panel enumerating the color of its node type. Here is an example of a multiple query, containing the drugs **"METHOTREXATE"** and **"INSULIN"**:

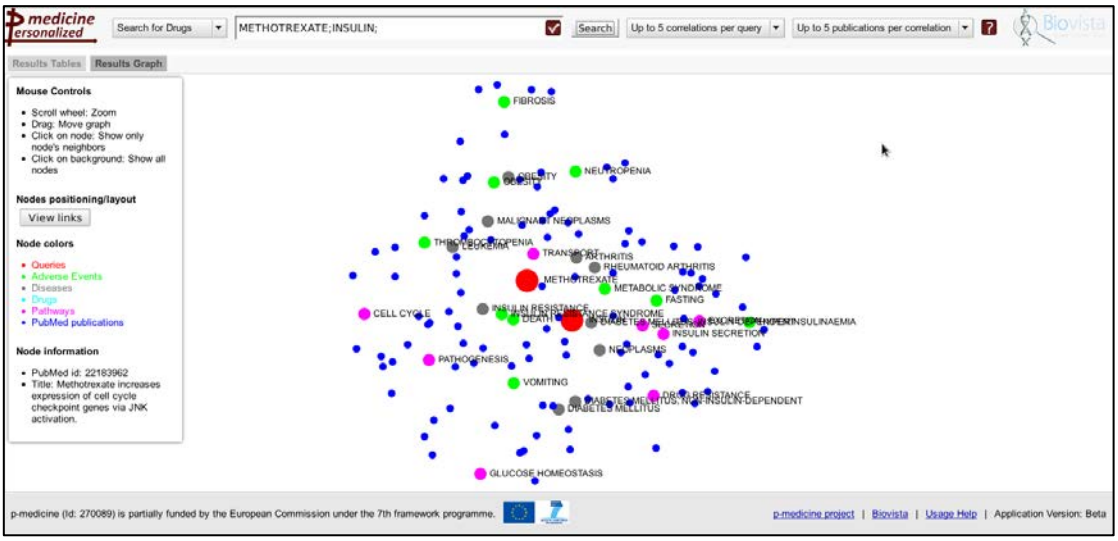


Figure 13: Graph overview

By default, after each performed search, the nodes are randomly placed on the two dimensional space and a positioning algorithm performs iterative automatic positioning in order to find the proper positions that reflect the node relationships. The algorithm calculates new positions by using the correlation scores and the common links between different nodes. If the user lets the algorithm to run multiple times, after a few seconds the positioning of nodes more accurately reflects their relationships:

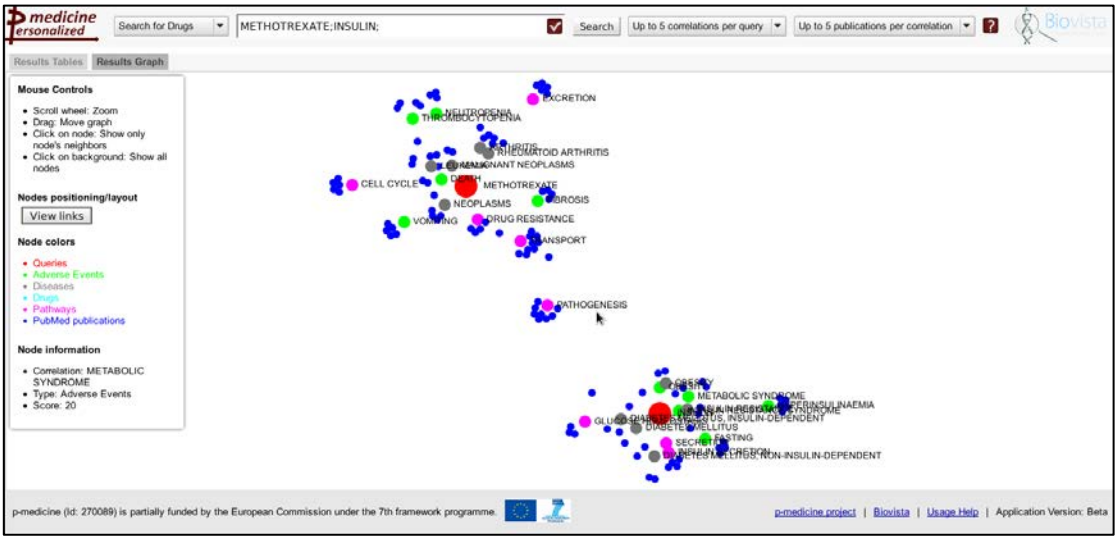


Figure 14: Graph auto-positioning

At any time, the user can stop the auto-positioning process and view the links between nodes by clicking on the **"View Links"** button on the left:

Figure 15: Click to view links

If the positions are not optimal, the auto-positioning process can be restarted by clicking again the button:

Figure 16: Click to restart auto-positioning

When the **"View Links"** button is pressed, all nodes freeze their positions and their relationships are displayed as lines connecting nodes:

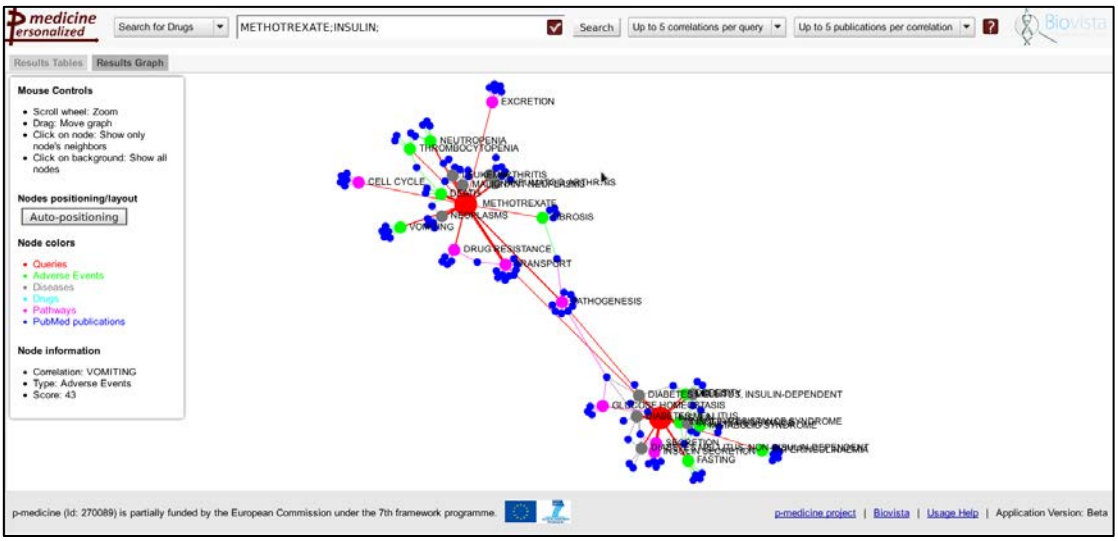


Figure 17: Graph links display

This view provides a quick overview of the result set and can be manipulated with the mouse. Scrolling the mouse wheel up, increases the zoom level of the graph area under the mouse cursor, scrolling down decreases it. This is especially useful for dense areas, where a large number of nodes are positioned close to each other. Clicking and dragging the mouse anywhere on the graph, moves the graph in the direction of mouse movement (panning). Clicking on a node, hides all nodes not linked to it, providing a quick overview of the nodes graph neighbors. To display all nodes again just click on the background, outside of any node. To quickly display information of a specific node please move the mouse cursor on that node and look at the left legend panel, under **"Node information"** section:

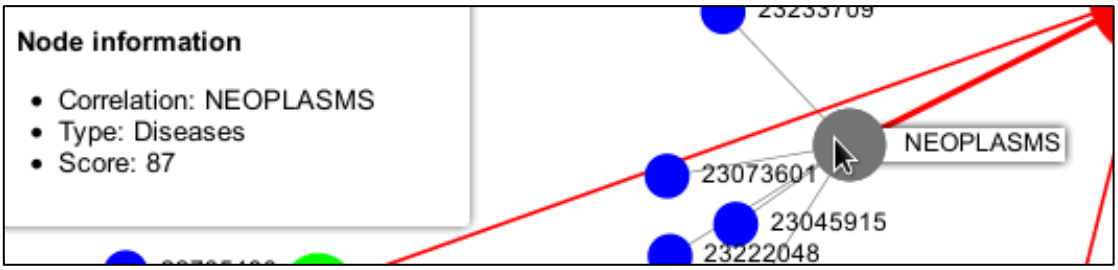


Figure 18: Node information

Publication nodes have an extra feature: moving the mouse cursor on them, a link to their corresponding page on the NCBI PubMed portal is popped up, so the user can direct his work to the full publication details:

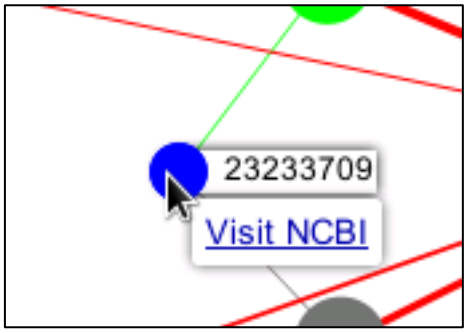


Figure 19: NCBI PubMed link

4 Literature Based Discovery Scenario

Clinical scenario I (Biovista)

4.1.1 Short description

Dabigatran etexilate is a new generation anticoagulant, operating as a direct thrombin inhibitor. Per the usual procedure, the drug has been tested in Phase III clinical trials prior to approval, where its Benefit/Risk (B/R) profile has been determined. However, once drugs go on the market, they are usually administered to a larger and much more variable patient population than the one studied in clinical trials. This is the reason why new Adverse Reactions (ADRs) are seen post-approval and are studied in Phase IV trial.

One important question regarding new generation anticoagulants, including dabigatran, is how their B/R profiles compare to those of existing anticoagulants, such as warfarin (Coumadin ®) and acenocoumarol (Sintrom ®). Here, we attempt to expand our understanding of the B/R profile of Dabigatran, and to identify patient subpopulations that are at risk for developing new ADRs, if any.

4.1.2 Steps for creating the scenario (biological/clinical)

The discovery process in the Biovista platform typically starts with defining one or more drugs of interest, here dabigatran etexilate. The first step is to understand the mechanism of action of the drug, which entails the identification of its main target and of its off-targets effects, if any.

In a second step, the pathways to which these targets belong are identified and characterized, in terms of:

- the complete sequence of proteins that they include
- their interactions with other known biological processes

Starting with dabigatran etexilate in the p-medicine correlation viewer, one immediately sees the top 15 correlated pathways in the biomedical literature (table 1). Please note that the 15 limit is imposed by the current version of the viewer, for performance reasons. More that 15 correlations can be returned by the COSS ™ Biovista Platform.

COAGULATION	AGING	FIBRINOLYSIS
BLOOD COAGULATION	PATHOGENESIS	HYPERSENSITIVITY
HEMOSTASIS	INTESTINAL ABSORPTION	GASTRIC MOTILITY
EXCRETION	ANGIOGENESIS	SECRETION
PLATELET AGGREGATION	CELL PROLIFERATION	PLATELET ACTIVATION

Table 1 – Top 15 pathways correlated to dabigatran etexilate, in the biomedical literature

It is interesting to note that although dabigatran is a thrombin inhibitor, and is thus understandably correlated in the literature to coagulation, hemostasis, platelet aggregation, platelet activation and fibrinolysis, it is also correlated to other, seemingly disparate, pathways, such as aging, angiogenesis, cell proliferation and gastric motility. This finding is a good starting point for exploring potential hitherto unknown ADRs of dabigatran and for identifying specific patient subpopulations that may be prone to the development of these ADRs.

Starting now with angiogensis, a pathway of particular interest, in a new p-medicine correlation viewer we get the following list of the top 15 highly correlated diseases (table 2):

TUMOR ANGIOGENESIS	CARCINOMA	HYPOXIA
NEOPLASMS	TUMOR GROWTH	MALIGNANT PARAGANGLIONIC NEOPLASM
PATHOLOGIC NEOVASCULARIZATION	CELL INVASION	MALIGNANT NEOPLASM OF BREAST
MALIGNANT NEOPLASMS	INFLAMMATION	ISCHEMIA
NEOPLASM METASTASIS	TISSUE ADHESIONS	MAMMARY NEOPLASMS

Table 2 – Top 15 diseases correlated to angiogenesis, in the biomedical literature

The list mainly includes neoplasms, inflammation and ischemic conditions; these certainly are results upon which one could expand, to determine what the B/R profile of dabigatran

could be in people also suffering from neoplastic, inflammatory or ischemic diseases. Using COSS ™ to further expand the list of diseases correlated to angiogenesis, we find at rank #43 wound healing. By studying the bibliography linking angiogenesis to wound healing, one understands that angiogenesis is an essential process, that must be successfully completed together with fibrosis, in order to form an evolving extracellular matrix and granulation tissue [1]. Dabigatran, on the other hand, has the potential to inhibit angiogenesis, through its induced inhibition of thrombin, while it has also been shown to exert antifibrotic effects in lung fibroblasts [2]. Therefore, dabigatran might impair wound healing. This hypothesis can now be looked into more carefully in future clinical trials; if proved correct, then the drug might need to be used with caution in patients with excessive wounds, such as those undergoing surgery or following trauma. The impairment of wound healing, if validated, comes as an additional complication, further to the excessive bleeding seen in wounded patients that are also treated with anticoagulants.

4.1.3 Perform scenario with Biovista's tools

- Correlation viewer/BEA and other tools of the platform are capable or returning direct relations between biological concepts. These can be used for exploring the correlation of dabigatran with biological pathways and diseases, as described above
- It is possible to access the supporting literature for a correlation in Biovista's tools. Analysis of the supporting evidence for the correlations discussed above can be carried out using this feature

4.1.4 References

1. Greaves NS, Ashcroft KJ, Baguneid M, Bayat A. Current understanding of molecular and cellular mechanisms in fibroplasia and angiogenesis during acute wound healing. J Dermatol Sci. 2013 Dec;72(3):206-17.

2. Bogatkevich GS, Ludwicka-Bradley A, Silver RM. Dabigatran, a direct thrombin inhibitor, demonstrates antifibrotic effects on lung fibroblasts. Arthritis Rheum. 2009 Nov;60(11):3455-64

Clinical scenario II (ecancer)

4.2.1 Short description

Clinicians are faced with a number of decisions such as which drug to prescribe on a daily basis. They cannot be aware of the latest research available on each of these topics as there is such a large amount of research being published, in 2004 Alper et al estimated it would take a physician trained in epidemiology 627.5 hours per month to keep up with the published research [1]. An online tool is therefore needed to aid healthcare professionals to search this literature and discover the results that are relevant to them as quickly as possible.

As a clinician faces a scenario such as the one as described above it is important they are aware of the benefits of using the Literature Based Discovery Platform instead of (or as well as) one of the other well-known online search facilities, such as Google, Bing etc. To ensure healthcare professionals alter their online search behaviour by using the Literature Based Discovery Platform, the online tutorial has 2 key objectives:

- Demonstrate clear additional to using the LBD over using existing online search platforms

- Ensure the healthcare professionals feel confident they can use the LBD effectively to deliver maximum benefit

To help achieve both of these goals, ecancer and Biovista have created a video tutorial that will be hosted on the ecancer.org platform. Clinicians will use this tutorial to support them as they use the tool to ensure they use if effectively to deliver the results the LBD is designed for and to ensure they receive significant additional benefit over using one of the standard online search tools.

4.2.2 Video tutorial development

Biovista created screencasts of the tool being used and provided a script explaining the processes being demonstrated. ecancer recorded a voiceover of the script and edited the clips together to form the first version of the tutorial video. This (version 1) was user tested on a number of clinicians and lay people to gather feedback on how well the tutorial delivered the stated objectives. The feedback received was that the language was 'too technical', therefore a second version of the script was written making the language more accessible for non-technical people, which the majority of clinicians are. This version (version 2) is currently available through ecancer's ftp site (http://share.ecancer.org/Pmed/Biovista/).

4.2.3 Future plans

The video tutorial will be further user tested on clinicians to ensure the objectives are being delivered. This user testing process will involve clinicians from multiple organisations across Europe. Once this testing process has taken place and the final video tutorial has been created it will be hosted on the personalised medicine area of the ecancer.org platform. This will be an area where users can quickly access all of the educational material created in the p-medicine project as well as accessing content such as news, videos and educational materials relating to the field of personalised medicine.

4.2.4 References

Alper BS, Hand JA, Elliott SG, Kinkade S, Hauan MJ, et al. (2004) How much effort is needed to keep up with the literature relevant for primary care? J Med Libr Assoc 92: 429–437.