Open PHACTS

Deliverable 5.1.4

First release of Target Dossier to community

Prepared by CNIO
Approved by PSMAR, AZ, Janssen, UNIVIE

November 2013
Version 1.0

Project title: An open, integrated and sustainable chemistry, biology and pharmacology knowledge resource for drug discovery
Instrument: IMI JU
Contract no: 115191

Start date: 01 March 2011
Duration: 3 years

Nature of the Deliverable	
Report	**x**
Prototype	
Other	
Dissemination level	
Public dissemination level	**x**
For internal use only	

Definitions

- Partners of the Open PHACTS Consortium are referred to herein according to the following codes:

  Pfizer – Pfizer limited – **Coordinator**
  UNIVIE – Universität Wien – **Managing entity of IMI JU funding**
  DTU – Technical University of Denmark – DTU
  UHAM – University of Hamburg, Center for Bioinformatics
  BIT – BioSolveIT GmbH
  PSMAR – Consorci Mar Parc de Salut de Barcelona
  LUMC – Leiden University Medical Centre
  RSC – Royal Society of Chemistry
  VUA – Vrije Universiteit Amsterdam
  CNIO – Spanish National Cancer Research Centre
  UNIMAN – University of Manchester
  UM – University of Maastricht
  ACK – ACKnowledge
  USC – University of Santiago de Compostela
  UBO – Rheinische Friedrich-Wilhelms-Universität Bonn
  AZ – AstraZeneca
  GSK – GlaxoSmithKline
  Esteve – Laboratorios del Dr. Esteve, S.A.
  Novartis – Novartis
  ME – Merck Serono
  HLU – H. Lundbeck A/S
  E.Lilly – Eli Lilly
  NBIC – Stichting Netherlands Bioinformatics Centre
  SIB – Swiss Institute of Bioinformatics
  ConnDisc – Connected Discovery
  EBI – European Bioinformatics Institute
  Janssen – Janssen Pharmaceutica
  OGL – OpenLink Software

- **Grant Agreement**: The agreement signed between the beneficiaries and the IMI JU for the undertaking of the Open PHACTS project.
- **Project**: The sum of all activities carried out in the framework of the Grant Agreement.
- **Work plan**: Schedule of tasks, deliverables, efforts, dates and responsibilities corresponding to the work to be carried, out as specified in the Grant Agreement.
- **Consortium**: The Open PHACTS Consortium composed of the above-mentioned legal entities.
- **Project Agreement**: Agreement concluded amongst Open PHACTS participants for the implementation of the Grant Agreement. Such an agreement shall not affect the parties' obligations to the Community and/or to one another arising from the Grant Agreement.

1 Target Dossier App

The Target Dossier (TD) goal is to provide a comprehensive view of pharmacologically relevant targets that allow answering questions regarding their druggability, tissue expression profiles and implications in pathways, disease states and physiological mechanisms.

As the name indicates TD is mainly focused on drug targets allowing users to explore target relevant data gathered from very different resources. It could be argued that several systems and databases sharing the same goal already exist (see for example (Gaulton et al., 2012) (Wu et al., 2009)), however, the complexity of this integration process justifies the co-existence of different tools/approaches. Particularly this integration entails two main challenges: the information technology challenge (how to combine the data in term of software, APIs, etc.) and the scientific challenge (which data should be taken into account and which methodologies must be followed to integrate the data and extract new knowledge).

Despite the fact that TD meets the requirements for open-source software products developed in the academia, the major goal during the TD development has been to find meaningful ways to combine bioactivity data with resources describing the target's functional role in the cell.

1.1 Release of the TD app v1.0

TD app v1.0 is a Web application built using state of the art standard compliant technologies. The application implements the MVC software design pattern and is mainly coded in Ruby on Rails and JavaScript. Sencha ExtJS (http://www.sencha.com/products/extjs/), an advanced client MVC JavaScript framework, has been used for the construction of the views and components. ExtJS supports modern Web browsers and offer a smoothly interactive experience to the users. TD application can be deployed in almost any system that has Ruby installed just by following the standard procedures for the deployment of Ruby on Rails applications. The version 1.0 is currently hosted in CNIO and is freely available at: http://td.inab.org.

Since the initial stages of the TD development it was clear the need of retrieve data not only from the OPS core API but also from the well-established collection of web services already available within the bioinformatics community (for example the IntAct database web services). The main reason behind this decision is to maintain the backward compatibility with some historical resources that doesn't provide a simple way to transform their data into RDF files. The dual retrieval system conceived and implemented to handle this situation is represented in figure 1.

The TD dual retrieval system distributes the queries among the TD engine and the OPS core API. While the OPS core API provides access to the data stored in the OPS linked cache the TD engine provides direct access to several web services already available in Internet. The dual retrieval system is a Ruby controller that gathers the responses from both sides and

sends a unified response to the TD frontend formed by the ExtJS components. This operation is completely transparent to the user who receives the information in real time.

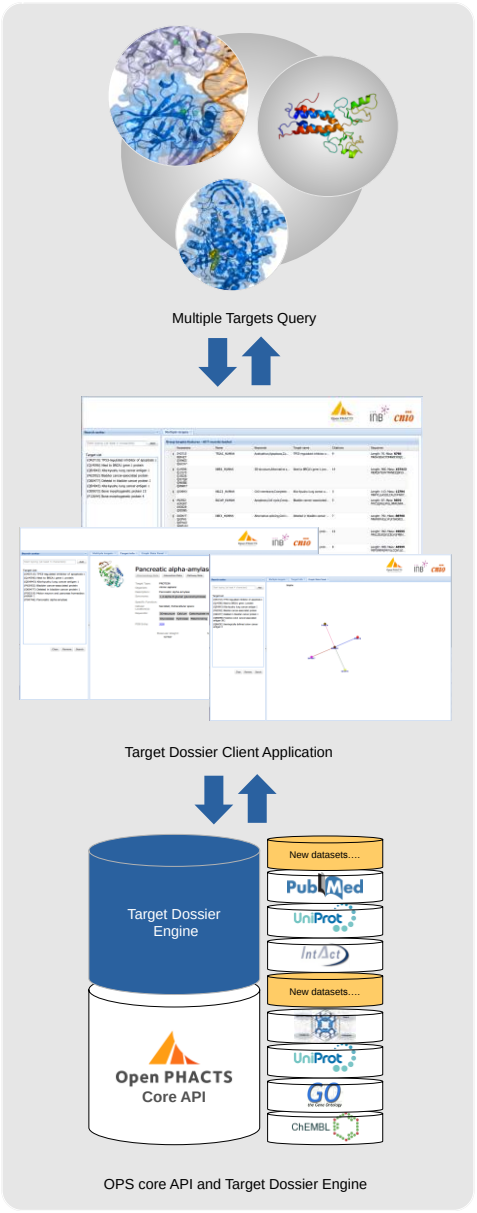

Figure 1. Target Dossier dual retrieval system. Queries are routed in parallel through the TD engine and the OPS API.

1.2 Use case scenarios

The TD app essentially provides tools to satisfy two general use cases. In the first case users can discover targets by entering keywords in the semantic search engine, composing a list of targets and play around in order to get relationships among them and with possible drugs. This scenario corresponds to some of the business cases defined by the WP6 and listed in the deliverable D6.1. For example:

- (Q44) - Give me all active compounds on a given target with the relevant assay data
- **(Q59)** - Identify all known protein/protein interaction inhibitors

The second case scenario implies to use the Hypothesis Tester that will be described below.

1.2.1 Semantic search and multiple target lists

The first use case starts with a semantic search. When the application is started up, a list of example targets is loaded by default. New targets can be searched using the semantic search box (Figure 2). When writing a text on the text field, a search is performed based on the input concept, accessing both OPS ConceptWiki (for targets) and Uniprot to retrieve and merge (if necessary) the data found. In such a way, several targets related to the same concept can be added to the list to compound a list of related targets.

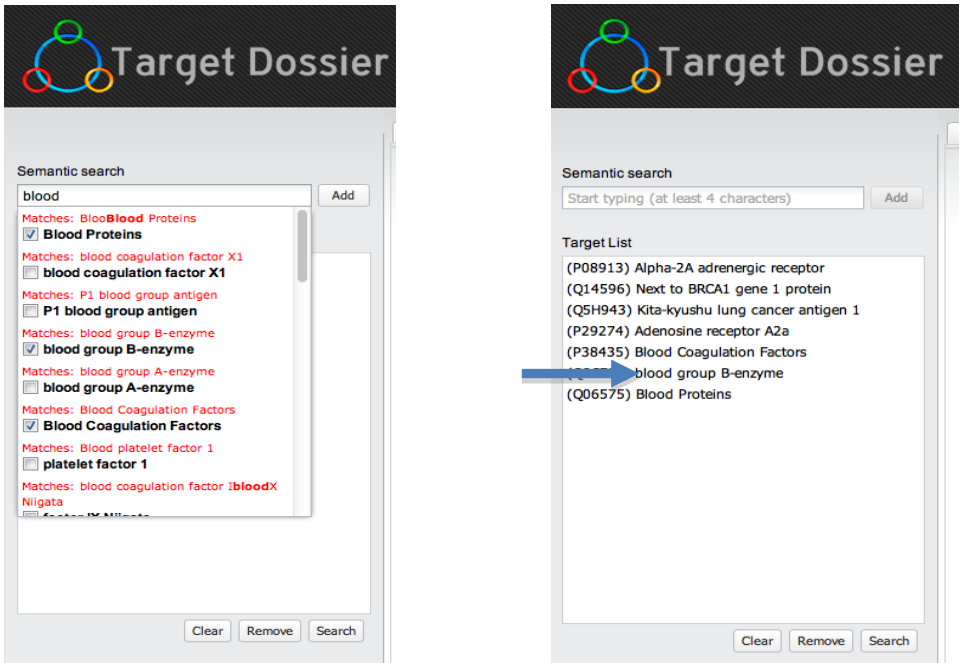

Figure 2: Semantic search and results selection (left). Then, the selected targets are added to the targets list after clicking the Add button.

From the selected list of targets (Figure 2, left), a summary report list (Figure 3) is displayed for the targets. More detailed information is obtained by double-clicking on any item to display a Target Information tab (Figure 4). From this point, interactions and drug related information could be obtained for the targets.

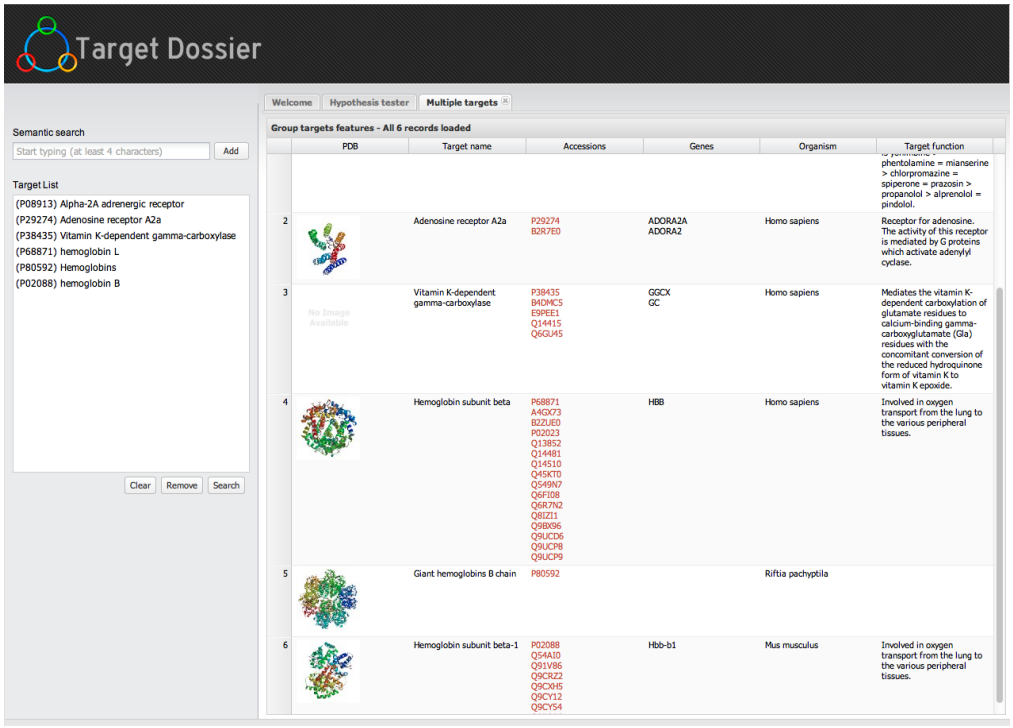

Figure 3. Summary information about the targets. This is displayed after clicking the '*Search*' button just below the target list. Each item in the list displays the target information by double-clicking it (Figure 4).

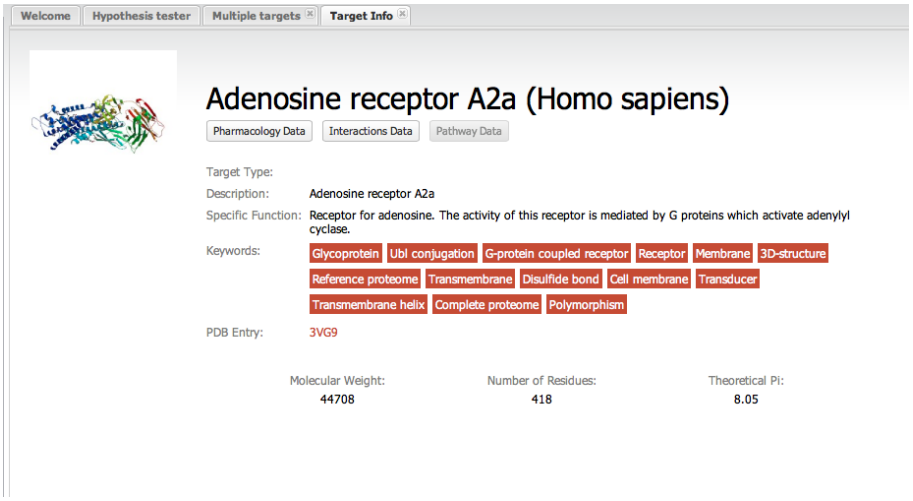

Figure 4. Target information tab, showing all information available for a target (full information in this case). By clicking the 'Pharmacology' or 'Interaction' buttons it is possible to get further information for the target, if available (Figure 5 and 6).

By discovering target interactions, new targets related to the chosen one can be added to the current list of targets in an iterative process (Figure 5). From the graph, target-target interaction information provided by IntAct and target information from OPS/Uniprot is displayed. By hovering on the nodes, information about the interactor (target) is displayed along with the chance to add that interactor to the target list.

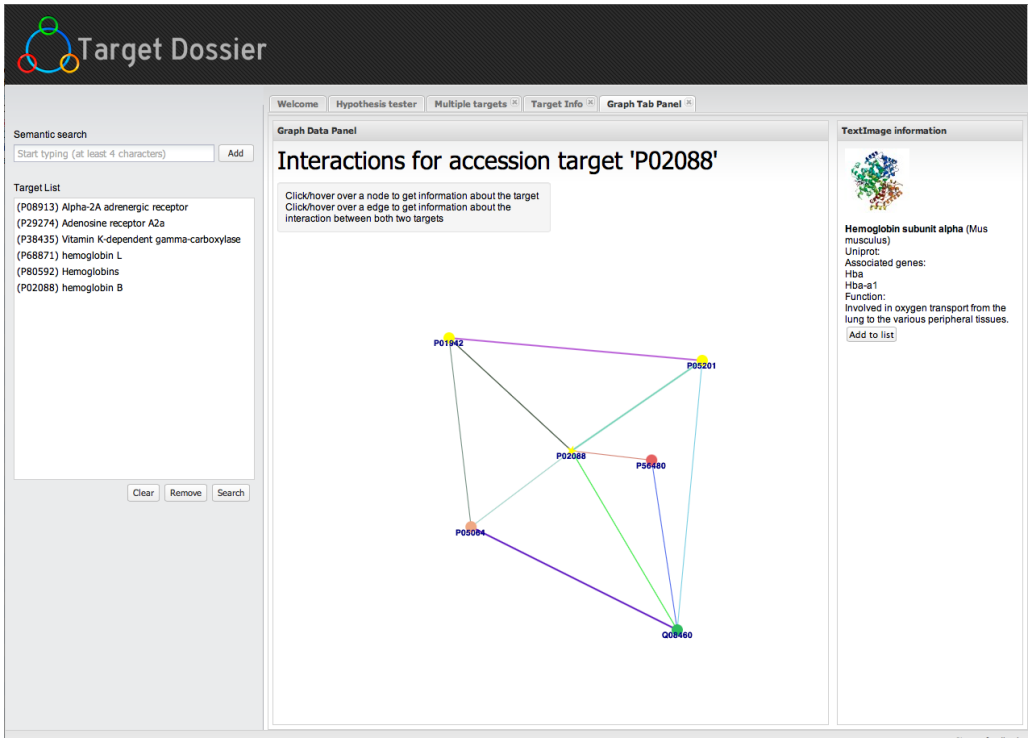

Figure 5: Interaction network. The point in the centre of the graph is the target that the interaction network was requested for. The width of the edges represents the number of interaction experiments found for the targets the edge joins (the more interactions found, the wider the edge will be). Hovering on both edges and nodes shows information about the interaction or the interactor on the right side (in this case showing information for a target).

Hovering on the edges displays the experiments which that interaction was found in, as well as external links to the IntAct website in the case of further information about the experiment is demanded.

Pharmacology data for a target is obtained from OPS platform and displayed in a table, showing a set of compounds related to the target, as showed in Figure 6. It displays complete information about the compounds related or which interact with the target (?)

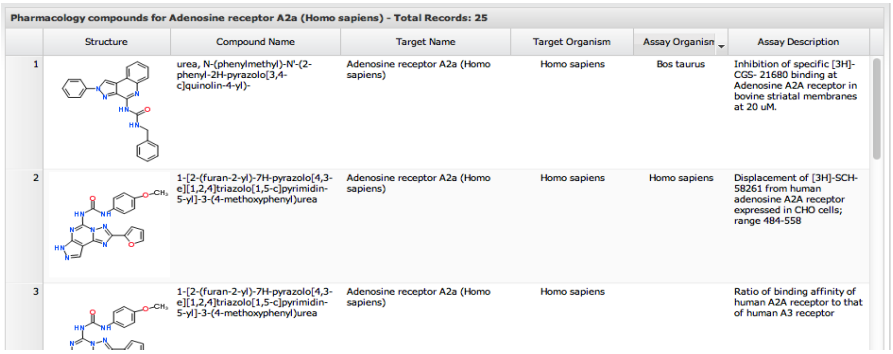

Figure 6: List of compounds for a target with full information for each compound

1.2.2 Hypothesis tester

This is another strategy to deal with the problem of discovering new drugs-target relationships and, in some way, represents an approach closer to the needs of the pharma

industry. Basically, the Hypothesis Tester (HT) tests whether a hypothesis is true by applying a set of programmatic rules. The hypothesis can be outlined as 'would like to know if aciclovir (a compound) has any effect on herpes' or for example the gene HER2 (Human Epidermal Growth Factor Receptor 2), has a major role in breast cancer'. Even if these questions might sound trivial (once known the answers) to experts they might serve to illustrate the principle. HT can answer these hypothesis and questions by digging in different data-sources as well as using the OPS API and applying different criteria on the retrieved data.

The starting point is, as in the case of the multiple targets use scenario, a semantic search. The previously described search tried to find only targets, but HT involves more biological entities (Figure 6) and then several data-sources are used to get results for any of the four types of entities (*Protein*, *Compound*, *Disease*, *Gene*).

When choosing a result, by clicking the side 'Add' button, the chosen concept is added to the graph playground (Figure 6). The point here is adding entities and setting up the hypothesis by joining the entities. As it is showed in Figure 7, several hypotheses can be 'chained' to build up a directed graph and discover novel relationships through intermediate steps.

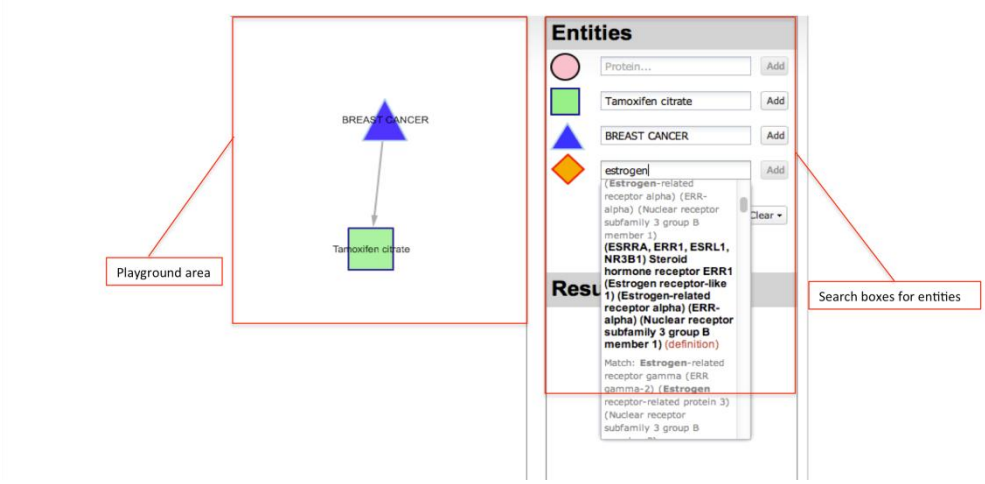

Figure 6. Playground for defining hypotheses as a graph (left) along with the semantic search for the entities (right). After choosing a result from the search, it can be added to the playground by clicking the 'Add' button.

The limitations for this approach come from the quantity and quality of the data in the external data-sources and the services and data itself exposed by them as an API. The tool does not attempt to make use of local data-sources if possible (actually the only exception is IntAct, as the performance of querying it was very low) as the data-sources which are queried (like OMIM) must have better data than we can come up with by ourselves.

Figure 7 shows a graph representing the molecular targets of the drug Lapatinib when used to treat breast cancer. As expected the green arrows indicate that an evidence exist confirming the relation between the disease breast cancer and the genes EGFR and HER2. At the same time the data validate those genes as a Lapatinib targets.

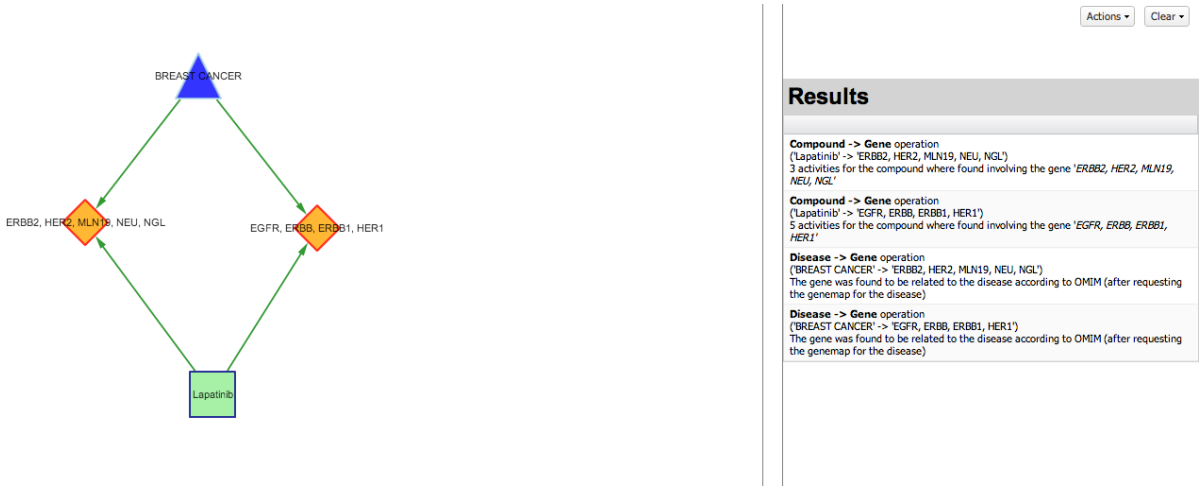

Figure 7. Lapatinib, breast cancer and the related oncogenes.

Data included in this release containing target information

- OPS
    a. ConceptWiki (http://www.conceptwiki.org/)
    b. ChemSpider (http://www.chemspider.com/)
    c. ChEMBL (http://www.ebi.ac.uk/chembl/)
- TD Engine
    a. PubMed (http://www.ncbi.nlm.nih.gov/pubmed)
    b. UniProt (http://www.uniprot.org/)
    c. IntAct (http://www.ebi.ac.uk/intact/)

2 References

Ecker, G. F., Kidd, R., & Williams-jones, B. (n.d.). Knowledge Driven Drug Discovery goes Semantic, 39–43.

Gaulton, A., Bellis, L. J., Bento, a P., Chambers, J., Davies, M., Hersey, A., Light, Y., et al. (2012). ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic acids research*, *40*(Database issue), D1100–7. doi:10.1093/nar/gkr777

Giardine, B., Riemer, C., Hardison, R. C., Burhans, R., Shah, P., Zhang, Y., Blankenberg, D., et al. (2005). Galaxy : A platform for interactive large-scale genome analysis, 1451–1455. doi:10.1101/gr.4086505.

Hopkins, A. L. (2008). Network pharmacology: the next paradigm in drug discovery. *Nature chemical biology*, *4*(11), 682–90. doi:10.1038/nchembio.118

Hudes, G., Carducci, M., Tomczak, P., Dutcher, J., Figlin, R., Kapoor, A., Staroslawska, E., et al. (2007). Temsirolimus, interferon alfa, or both for advanced renal-cell carcinoma. *The New England journal of medicine*, *356*(22), 2271–81. doi:10.1056/NEJMoa066838

Hull, D., Wolstencroft, K., Stevens, R., Goble, C., Pocock, M. R., Li, P., & Oinn, T. (2006). Taverna: a tool for building and running workflows of services. *Nucleic acids research*, *34*(Web Server issue), W729–32. doi:10.1093/nar/gkl320

Lu, J.-J., Pan, W., Hu, Y.-J., & Wang, Y.-T. (2012). Multi-target drugs: the trend of drug research and development. *PloS one*, *7*(6), e40262. doi:10.1371/journal.pone.0040262

Luciano, J. S., Andersson, B., Batchelor, C., Bodenreider, O., Clark, T., Denney, C. K., Domarew, C., et al. (2011). The Translational Medicine Ontology and Knowledge Base: driving personalized medicine by bridging the gap between bench and bedside. *Journal of biomedical semantics*, *2 Suppl 2*(Suppl 2), S1. doi:10.1186/2041-1480-2-S2-S1

Motzer, R. J., Escudier, B., Oudard, S., Hutson, T. E., Porta, C., Bracarda, S., Grünwald, V., et al. (2008). Efficacy of everolimus in advanced renal cell carcinoma: a double-blind, randomised, placebo-controlled phase III trial. *Lancet*, *372*(9637), 449–56. doi:10.1016/S0140-6736(08)61039-9

Pujol, A., Mosca, R., Farrés, J., & Aloy, P. (2010). Unveiling the role of network and systems biology in drug discovery. *Trends in pharmacological sciences*, *31*(3), 115–23. doi:10.1016/j.tips.2009.11.006

Szklarczyk, D., Franceschini, A., Kuhn, M., Simonovic, M., Roth, A., Minguez, P., Doerks, T., et al. (2011). The STRING database in 2011: functional interaction networks of proteins, globally integrated and scored. *Nucleic acids research*, *39*(Database issue), D561–8. doi:10.1093/nar/gkq973

Wang, Y., Xiao, J., Suzek, T. O., Zhang, J., Wang, J., Zhou, Z., Han, L., et al. (2012). PubChem's BioAssay Database. *Nucleic acids research*, *40*(Database issue), D400–12. doi:10.1093/nar/gkr1132

Wu, C., Orozco, C., Boyer, J., Leglise, M., Goodale, J., Batalov, S., Hodge, C. L., et al. (2009). BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology*, *10*(11), R130. doi:10.1186/gb-2009-10-11-r130