

# Studying toxicogenomics data for phthalates using ToxWiz™

Gordana Apic<sup>1</sup>, Frank Bonner<sup>1</sup>, Bojana Cosovic<sup>1</sup>, Robert B. Russell<sup>2</sup>, Vanessa J. Hoy<sup>3</sup>

1. Cambridge Cell Networks Ltd., St John's Innovation Centre, Cowley Road, Cambridge, CB4 0WS, UK
2. EMBL, Meyerhofstrasse 1, 69117 Heidelberg, Germany
3. ScriboScience Ltd., 10 Laceys Lane, Exning, Newmarket, Suffolk, CB8 7HL, UK

In this case study we examine a mouse liver microarray dataset for the well established toxin, phthalate. We also investigate the phthalate molecule itself, using a substructure search, then compare the results with those that we get from the microarray.

## Introduction

This case study illustrates some of the basic features of ToxWiz™ for studying toxicogenomics data.

ToxWiz is a comprehensive software and database system designed to help make decisions in toxicology and many other areas of drug discovery.

The system works by exploiting a network of thousands of interacting genes, proteins, chemicals, pathways and biological effects, including toxic endpoints and diseases. The neighbourhood of a molecule in the system captures everything that has been observed about it in the past, either in the public domain or in company proprietary datasets.

Here we examine a mouse liver microarray dataset for the well established toxin, phthalate. We also investigate the chemical itself, by way of a substructure search, and compare the results with those that we get from the microarray.

## Importing the micro-array data

We begin by loading the microarray dataset which was extracted from the

Environment, Drugs and Gene Expression (EDGE) database. The dataset comprises a mouse liver microarray taken at one time point, for animals given diethylhexyl phthalate<sup>1</sup>.

ToxWiz is human centric but it also contains data for 17 other species including higher mammals such as chimp, dog and mouse, as well as simpler organisms like zebrafish, nematode and yeast. When animal data is imported ToxWiz determines how many of the genes in the animal it can equivalence in humans. Only genes about which something is known, either in terms of their

function or their interactions with proteins and chemicals, are reported. Typically, this is between 50% and 70% of the genes in the dataset.

It is also possible to specify thresholds for the genes to include in the analysis. Here we will only look at up regulated genes with a fold change of 1.5 or higher.

## Mapping the micro-array data onto toxic endpoints.

A quick and easy way to see what genes might do is to map them onto pathways or clusters in the system. A cluster is a set of genes or chemicals associated with a biological phenomenon such as a toxic endpoint or a disease. In this case we will consider only toxic endpoints (Figure 1).

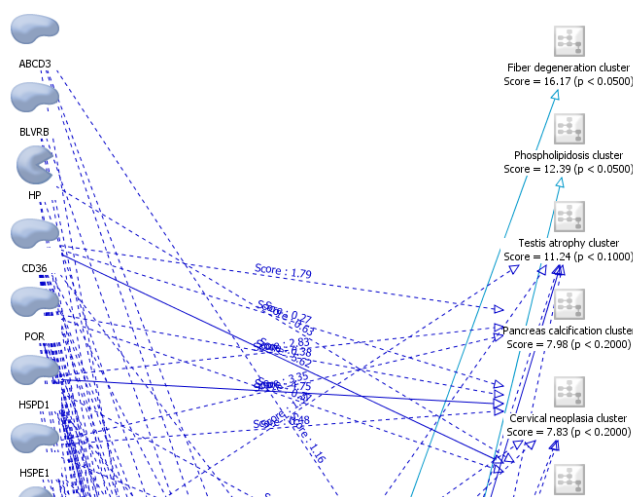
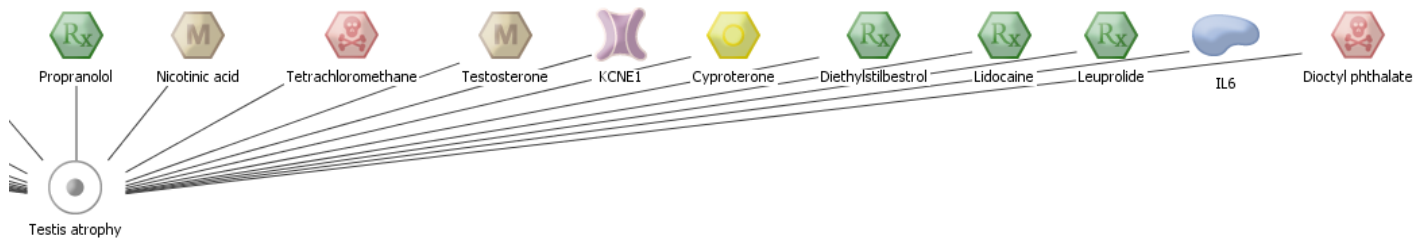
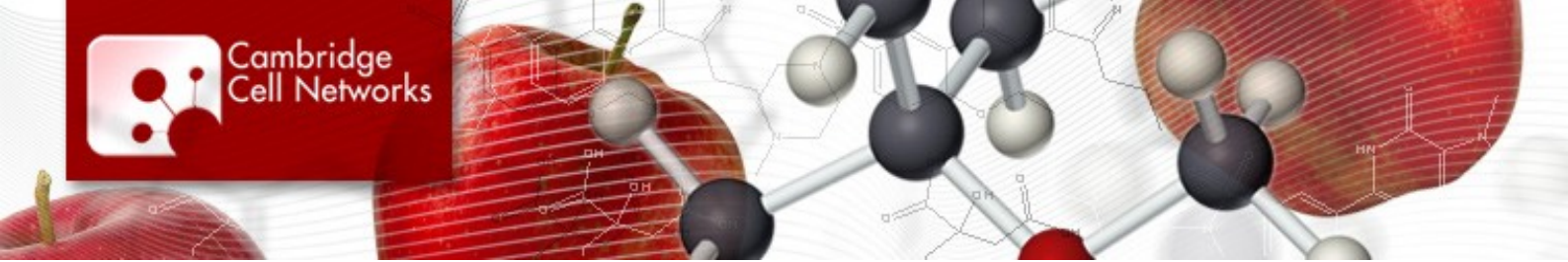


Figure 1 Mapping of mouse liver microarray data for diethylhexyl phthalate onto toxic endpoints



**Figure 2** Some of the contents of the testis atrophy toxic endpoint cluster

The map shows genes down the left side, connected to toxic endpoints on the right. Solid lines connecting genes to toxic endpoints indicate that the gene is inside the cluster. Dashed lines indicate that the gene is interacting or influencing something that is inside the cluster.

The clusters are ranked according to our statistical mapping system which considers how common genes are in the system, in addition to connections between genes.

The third ranked toxic endpoint in this case is testis atrophy. This is a well known problem with phthalate molecules, which cause genital and germline development problems.

### Exploring toxic endpoints

A toxic endpoint cluster is simply a set of genes, proteins and chemicals associated somehow with that endpoint. Figure 2 shows part of the

testis atrophy cluster. There are many items linked to the testis atrophy toxic endpoint, suggesting that this disease cluster has been extensively studied. The genes from our microarray dataset which are inside this cluster appear in blue, with bars indicating the degree of up or down regulation. Hovering over these bars displays the numbers from the original datafile.

We can also see how other genes in our dataset are indirectly connected to this cluster.

If we consider all types of connections, then several other genes from our dataset are added to the cluster (Figure 3). These connections contributed to the score that we got in our previous mapping (Figure 1). This is one of the powerful and unique features of ToxWiz.

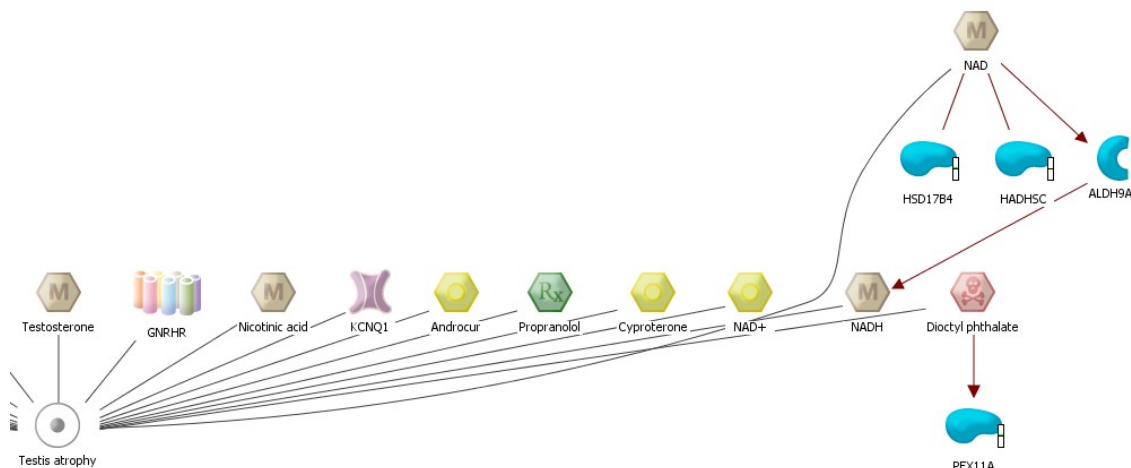
The expanded cluster reveals that phthalates have previously been

seen to up regulate PEX11A<sup>2</sup>. This gene is also up regulated in our dataset and whilst there is no evidence of a direct link between PEX11A and testis problems in the current literature, ToxWiz has inferred such a relationship, based on the vast network of genes, proteins and chemicals stored in the database.

Other genes which are known to have an effect on testis development, are also present in the cluster, such as cytochrome c, which causes problems when it is deleted in the mouse<sup>2</sup>.

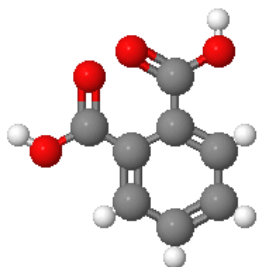
### Searching for phthalate compounds

Now we will investigate other phthalates in the database. A chemical substructure search for a phthalic acid fragment (Figure 4) inside a larger molecule yields 19 hits out of the roughly 20,000 chemicals



**Figure 3** Expanded testis atrophy toxic endpoint cluster showing connected genes from the mouse liver micro-array dataset

that are currently linked to proteins, genes or biological effects in the system.



**Figure 4** Phthalic Acid

Mapping these chemicals onto toxic endpoints gives good agreement with the microarray data. This is probably no surprise, as phthalates are well known to cause these problems.

### Combining the results of the substructure search with the micro-array data

However, combining this chemical information with our microarray dataset does provide further insights. Figure 5 shows part of the highest scoring cluster from the previous map, testis degeneration, with both the connected chemicals from the phthalate substructure search and the connected genes from the microarray dataset added.

This gives us a better picture of what might be happening when mice are

given these compounds. The figure shows that phthalates are known to activate or induce various nuclear receptors such as PPAR $\alpha$  and RAR $\alpha$ <sup>2</sup>. Moreover, several genes from our micro-array dataset appear to be activated by one or more of these transcription factors, such as the transporter molecule, ABCD3.

### Summary

This analysis demonstrates many of the ways in which ToxWiz can be used to investigate toxicogenomics data. It shows how by combining microarray data with the chemical information in the system it is possible to gain insights into possible mechanisms of toxicity.

### References

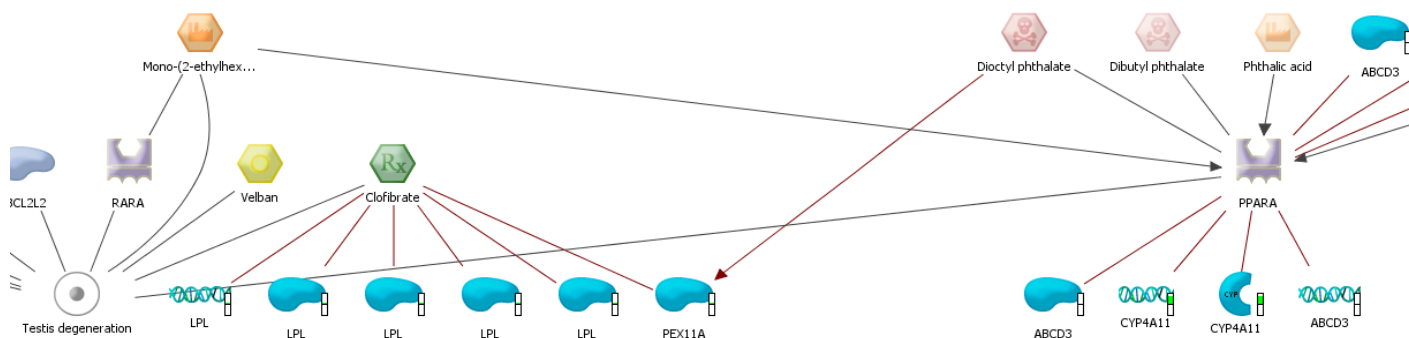
- 1.
2. Narisawa et al (2002) Mol Cell Biol **22**(15):5554-5562.

### About Cambridge Cell Networks

Cambridge Cell Networks (CCNet) based in Cambridge, UK supplies a range of industry-leading content on biological pathways, chemistry and toxicology, combined with an integrated pathway visualisation and exploration tools to the pharmaceutical and biotechnology industries. Using cutting-edge

biological and computational methods combined with knowledge management techniques, CCNet offers a novel approach to pathway analysis, providing effective target validation and predictive toxicology data, which will ensure the production of safer drugs.

The Company was founded in 2002 by eminent scientific figures from the European Bioinformatics Institute in Cambridge, UK, the University of Cambridge and the European Molecular Biology Laboratory in Heidelberg, Germany. CCNet has facilities in three countries and is staffed by a team of expert biochemists, pharmacologists, bioinformaticians, chemists and industrial toxicologists.



**Figure 5** Testis degeneration toxic endpoint cluster showing connected chemicals from the substructure search and connected genes from the mouse liver micro-array dataset