Literature-based compound profiling: application to toxicogenomics

Raoul Frijters[1], Stefan Verhoeven[2], Wynand Alkema[2], René van Schaik[1,2] & Jan Polman[2†]

†Author for correspondence
[1]Radboud University Nijmegen Medical Centre, Centre for Molecular and Biomolecular Informatics (CMBI), Nijmegen Centre for Molecular Life Sciences (NCMLS), PO Box 9101, 6500 HB Nijmegen, The Netherlands
[2]Organon NV, Department of Molecular Design & Informatics, PO Box 20, 5340 BH Oss, The Netherlands
Tel.: +31 412 663 678; Fax: +31 412 662 539; E-mail: jan.polman @organon.com

Introduction: To reduce continuously increasing costs in drug development, adverse effects of drugs need to be detected as early as possible in the process. In recent years, compound-induced gene expression profiling methodologies have been developed to assess compound toxicity, including Gene Ontology term and pathway over-representation analyses. The objective of this study was to introduce an additional approach, in which literature information is used for compound profiling to evaluate compound toxicity and mode of toxicity. **Methods:** Gene annotations were built by text mining in Medline abstracts for retrieval of co-publications between genes, pathology terms, biological processes and pathways. This literature information was used to generate compound-specific keyword fingerprints, representing over-represented keywords calculated in a set of regulated genes after compound administration. To see whether keyword fingerprints can be used for assessment of compound toxicity, we analyzed microarray data sets of rat liver treated with 11 hepatotoxicants. **Results:** Analysis of keyword fingerprints of two genotoxic carcinogens, two nongenotoxic carcinogens, two peroxisome proliferators and two randomly generated gene sets, showed that each compound produced a specific keyword fingerprint that correlated with the experimentally observed histopathological events induced by the individual compounds. By contrast, the random sets produced a flat aspecific keyword profile, indicating that the fingerprints induced by the compounds reflect biological events rather than random noise. A more detailed analysis of the keyword profiles of diethylhexylphthalate, dimethylnitrosamine and methapyrilene (MPy) showed that the differences in the keyword fingerprints of these three compounds are based upon known distinct modes of action. Visualization of MPy-linked keywords and MPy-induced genes in a literature network enabled us to construct a mode of toxicity proposal for MPy, which is in agreement with known effects of MPy in literature. **Conclusion:** Compound keyword fingerprinting based on information retrieved from literature is a powerful approach for compound profiling, allowing evaluation of compound toxicity and analysis of the mode of action.

Keywords: compound profiling, diethylhexylphthalate, dimethylnitrosamine, gene expression, literature-based networks, methapyrilene, microarray, profiling, text mining, toxicogenomics

In the drug development process it is important that any toxicity of new compounds is recognized as early as possible to guarantee drug safety and to reduce the costs of drug development [1]. Classical methods of toxicity evaluation of compounds involve elaborate and time-consuming animal experiments in which a number of measurements are carried out to evaluate toxicity, including clinical chemistry, hematology, histopathology and *in vitro* assays.

In recent years, toxicogenomics has emerged as a new methodology to evaluate compound toxicity. With toxicogenomics, compound-induced gene expression profiles are analyzed to detect indications of toxicity and, as such, may provide a promising way to rapidly screen for adverse drug effects [2,3]. The challenge in this approach is to infer toxicity of the compound from the gene expression profiles in order to predict the toxicity of a compound and to resolve the underlying mechanism of toxicity.

One approach relies on comparison of observed gene expression profiles with a database of expression profiles for reference compounds with known toxic effects [4].

Another approach consists of the analysis of the annotation of the genes that are differentially expressed in a drug-treated sample. Multiple algorithms exist that analyze gene sets for the presence of over-represented Gene Ontology (GO), Kyoto Encyclopedia of Genes and Genomes (KEGG) or other annotation terms [5,6]. These algorithms have been applied successfully to elucidate toxic effects and mode of action of a number of compounds [7].

In this paper we want to introduce an additional approach to link gene expression profiles to

10.2217/14622416.8.11.1521 © 2007 Future Medicine Ltd ISSN 1462-2416

toxicity. We developed a tool named CoPub that is able to calculate keyword over-representation in a similar fashion to general GO term over-representation tools but, in our approach, the annotation of the genes is retrieved directly from Medline by text mining [101]. Medline represents an enormous amount of information on the function of genes/proteins and their role in the onset of toxicity, and is therefore an ideal source to identify relations between the expression of genes and toxicity end points.

Several text mining methods for the analysis of microarray data have been published, including methods for the generation of literature neighborhoods from regulated genes [8] and clustering a set of regulated genes based on their literature profile [9–13]. Other methods are aimed at annotating gene sets that are obtained by clustering of genes based on their expression profile, often based on subsets of the total Medline repository [14,15]. CoPub uses the entire Medline library to calculate robust statistics for gene-keyword co-occurrence, and is not dependent on preclustered gene sets to calculate significance for keyword over-representation.

CoPub is built on the assumption that co-citation of a gene and a keyword in the same Medline abstract is indicative of a relation between the gene and the keyword and that the keyword thus represents some sort of annotation of the gene. Using thesaurus-based keyword matching [8,16,17] with thesauri for genes, biological processes, drugs, liver pathologies and diseases, all keyword co-publication hits in the Medline abstracts database are collected. This co-publication information is then used to calculate keyword over-representation in sets of genes that are differentially expressed in compound-treated samples. The list of over-represented keywords can be analyzed subsequently to assess the toxicity and mode of toxicity of the compounds that are tested.

To demonstrate the value of CoPub, we applied this text mining tool to two microarray data sets. In one experiment, male Wistar Hanover rats were dosed daily by gavage with four nongenotoxic carcinogens; methapyrilene (MPy), diethylstilbestrol (DES), Wy-14643 (Wy) and piperonylbutoxide (PBO), and four genotoxic carcinogens; 4-(methylnitrosamino)-1-(3-pyridyl)-1-butanone (NNK), 2-nitrofluorene (2-NF), dimethylnitrosamine (DMN) and aflatoxin B1 (AB1) [18,19]. In the other experiment, Sprague-Dawley rats were given a single dose by oral administration with three peroxisome proliferators: clofibrate (Clo), valproic acid (VPA) and diethylhexylphthalate (DEHP) [20]. Histological examination of the livers treated with the four genotoxic carcinogens showed increasing necrosis after DMN and AB1 treatment, which was accompanied by reactive inflammation in both cases [19]. 2-NF induced modest hypertrophy and NNK showed only weak apoptosis after 14 days of treatment [19]. On a histopathological level, all four nongenotoxic carcinogens were shown to induce weak (PBO) to moderate (DES, Wy, MPy) hyperplasia and hypertrophy. MPy induced apoptosis with increasing severity over time, causing weak inflammation [18]. Histopathological examination of the livers treated with the three peroxisome proliferators Clo, DEHP and VPA showed some vacuolation, hypertrophy and increased mitotic figures [20].

Using CoPub, we generated keyword fingerprints for each of the compounds. Detailed examination of the regulated gene sets and their over-represented keywords for a genotoxic carcinogen, a nongenotoxic carcinogen and a peroxisome proliferator revealed mechanisms of toxicity for these compounds that correlate well with histopathological data. These results show that text mining of gene expression data, as implemented in CoPub, is a valuable tool for early screening of compounds for toxic effects.

Material & methods

Medline search strings definitions

Four thesauri were generated to search Medline:

- Genes (human, mouse, rat)
- Liver pathologies
- Biological processes
- Pathways

The keyword thesauri are based on biological items, which represent an instance of a biological concept (e.g., a gene, a pathway), and may contain one or more keywords (e.g., a gene is assigned a full gene name and/or a gene symbol).

Human, mouse and rat gene thesauri were compiled from the National Center for Biotechnology Information's (NCBI's) Entrez Gene database (released December 2005) [21,102]. In order to search Medline with one or more full gene names, gene symbols and aliases, the gene name thesauri were processed as described by Alako *et al.* [16]. Gene names and symbols of orthologous genes were combined to make the keyword search in Medline more comprehensive.

The pathology thesaurus contains 406 pathology terms and was compiled from textbooks and

is particularly focused on liver-specific pathologies (e.g., cholestasis), but also contains less specific liver pathologies (e.g., necrosis).

The GO biological process thesaurus was compiled from the GO database [103] and contains 5515 terms.

The pathway thesaurus contains 817 pathway names, compiled from the KEGG database [104], the encyclopedia of human genes and metabolism database [105] and the Reactome database [106].

The full Medline baseline XML files (between 1966 and April 2006) were obtained from the NCBI website [107] and extracted to small text files containing title, abstract and substances.

Regular expressions were used to search the compiled Medline text files for the presence of all keywords (~250,000) from the biological concept thesauri, as described by Alako *et al.* [16]. Keywords that generated a hit in a Medline abstract were stored, together with the PubMed identifiers (IDs) of the Medline records in which the hit occurred. For every biological item the hits were made nonredundant (note: multiple keywords of a biological item can occur in the same Medline abstract), resulting in a PubMed ID–biological item list. Gene symbols were curated for ambiguity [Frijters *et al.*, manuscript in preparation].

Co-publication of biological items (e.g., a gene with a pathology term) was retrieved from the database by matching common Medline abstract occurrences. An R-scaled score, described by Alako *et al.* [16], that describes the strength of a co-citation between two keywords given their individual frequencies of occurrence and the number of co-publications between every biological item pair was calculated in order to assign a degree of relation between two keywords.

The R-scaled score is based on the mutual information measure and was calculated as:

$$S = P_{AB} / P_A \times P_B$$

in which P_A is the number of hits for biological item A divided by the total number of PubMed IDs, P_B is the number of hits for biological item B divided by the total number of PubMed IDs, and P_{AB} is the number of co-occurrences between biological item A and biological item B divided by the total number of PubMed IDs. The relative score is produced as a 1–100 scaled $\log_{10}$ conversion ($R = {}^{10}\log S$) and the scaled-log-transformed relative score (R-scaled score) as:

$$R' = 1 + 99 \times (R - R_{min}) / (R_{max} - R_{min})$$

here R_{min} and R_{max} are the lowest and highest R values present in the biological item co-publication list, respectively.

Publicly available microarray data sets from hepatotoxicity studies on genotoxic and nongenotoxic carcinogens, and peroxisome proliferators were downloaded from the EBI (European Bioinformatics Institute) ArrayExpress database (E-TOXM-16 and E-TOXM-19 datasets) [22,108].

The downloaded .CEL files were imported in Rosetta Resolver v5.1 (Rosetta, Seattle, WA, US), and differentially expressed genes (p-value ≤ 0.01, fold-change ≥2 or ≤-2) were selected with the implemented Rosetta Resolver error-model using the corresponding time-matched controls. Control gene sets were generated by randomly selecting two sets of respectively 100, 150, 200, 250 and 300 Affymetrix probe set identifiers from the RG-U34A GeneChip array.

Links from Affymetrix IDs to Entrez Gene IDs and orthology information was retrieved from Affymetrix human, mouse, rat GeneChip Genome Array annotation files (sep. 2005) [109].

To determine keywords that are over-represented in a given gene set, all keywords that were associated with a gene (defined as having at least five co-citations) were collected. The same was done for the background set of genes consisting of nondifferentially expressed genes on the array. The degree of association between the gene set and a keyword was then determined using the Fisher Exact Test. Keywords associated with at least four genes in the gene set and with a p-value not more than 0.01 were marked as over-represented in the gene set after multiple test correction using the Benjamini-Hochberg correction, as implemented in the *R* Statistics package [110]. These settings were empirically determined following analyzed microarray data sets with known biological outcome (data not shown). We did not make a distinction between down- and upregulated genes in the regulated gene sets, because *a priori* it cannot be established whether upregulation or downregulation of a gene will contribute to the toxicity.

All statistical tests were done using the *R* Statistics package [110]. The literature network between over-represented keywords and genes (nodes) and their co-publications (edges) were visualized using Cytoscape software [23,111].

Results

Compound keyword fingerprints

To see whether toxicity of a compound can be assessed from the literature by keywords associated with the compound's gene expression profile, two

microarray data sets, including 11 hepatotoxicants, were analyzed. One data set was generated by treatment of rat liver with eight hepatocarcinogens [18] and the other by treatment of rat liver with three peroxisome proliferators [20]. The 11 compounds were analyzed for gene expression profiling and included four nongenotoxic hepatocarcinogens: MPy, DES, Wy and PBO; four genotoxic hepatocarcinogens: 2-NF, DMN, NNK and AB1; and three peroxisome proliferators: Clo, VPA and DEHP. It should be noted that Wy is also known as a potent peroxisome proliferator, and that peroxisome proliferators, if chronically administered, are carcinogenic (nongenotoxic) in rodents.

We identified sets of regulated genes for each of the 11 compounds using the raw data at the time point on which the compounds revealed their effects on a histopathological level; for the eight carcinogenic compounds this was after 7 days of compound administration, and for the three peroxisome proliferators after 2 days of compound administration. These gene sets were then analyzed for keyword over-representation, generating a keyword fingerprint for each compound (the keyword fingerprints are available in the online web supplement).

To study whether keyword fingerprints can be used to discriminate between compounds with different modes of action, we compared keyword fingerprints of two genotoxic carcinogens (DMN and AB1), two nongenotoxic carcinogens (MPy and PBO), two peroxisome proliferators (DEHP and VPA) and of two randomly generated gene sets (150 and 250 randomly selected Affymetrix probe set identifiers). For ease of comparison the keywords were categorized into 13 groups: pathogenesis, drug metabolism, oxidative stress, DNA damage and response, cell cycle and mitosis, apoptosis and cell death, inflammation and immune response, cell differentiation and development, steroid metabolism, lipid metabolism, energy metabolism, general cell metabolism, and miscellaneous keywords. The profiles of the keyword fingerprints are visualized in Figure 1 (the categories and their corresponding keywords can be retrieved from the online web supplement).

Notable differences in the keyword profiles can be appreciated in Figure 1. The keywords of the two peroxisome proliferators DEHP and VPA are mainly grouped into the steroid, lipid, and energy metabolism categories. The keyword fingerprints of DMN and AB1 strongly reflect processes of cell death, accompanied by inflammation, a hallmark of most genotoxic carcinogens, whereas oxidative stress, found with many nongenotoxic carcinogens, is found predominantly with PBO and MPy. Some categories are hit by all compounds, such as drug metabolism, general cell metabolism and, to a lesser extent, pathogenesis. Similar profiles were observed for the other compounds, although these profiles were less pronounced due to the fact that these compounds induced a smaller number of genes. All of the randomly generated gene sets produced flat profiles with no keyword enrichment, indicating that the keyword fingerprints of the analyzed hepatotoxicants reflect real biological processes rather than random noise.

These results show that literature associations between genes and keywords can be used to discriminate between compounds with distinct biological activities.

Comparing keyword fingerprints

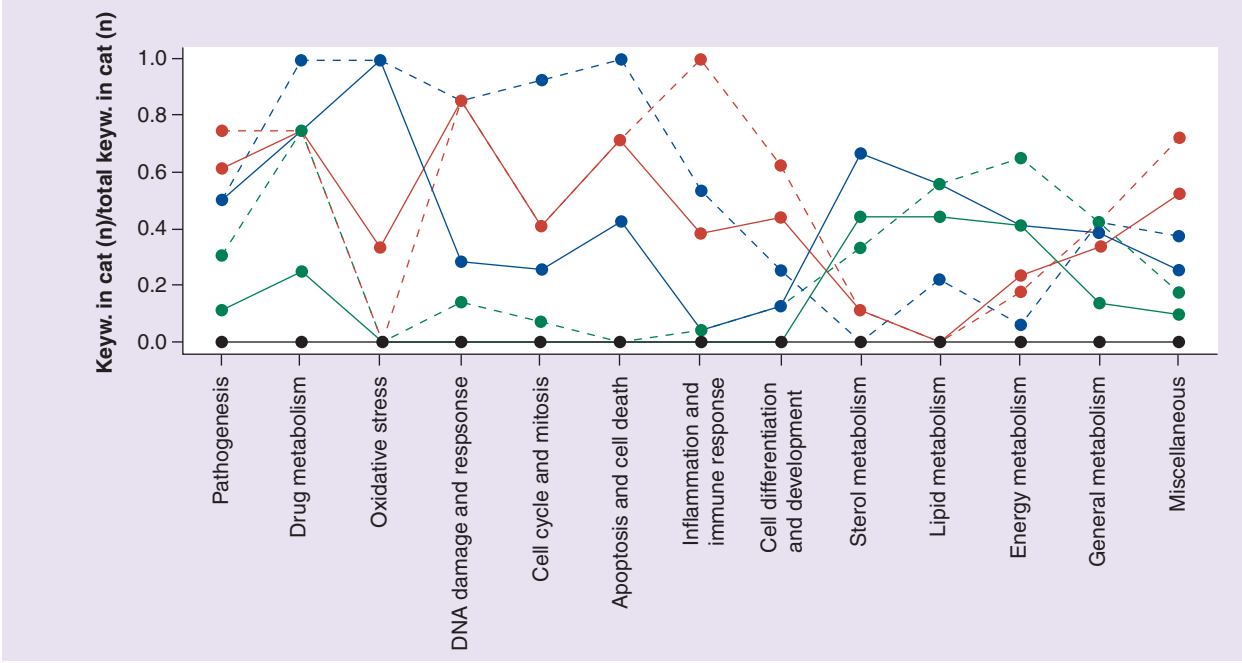
To see if the observed distinct fingerprints between compounds reflects the biological and toxicological effects of the compounds, we analyzed the keyword fingerprints of a peroxisome proliferator (DEHP), a genotoxic carcinogen (DMN) and a nongenotoxic carcinogen (MPy), in more detail (Table 1; Note: To reduce the size of Table 1 we do not show p-values. Compound-specific keyword fingerprints, representing over-represented keywords with their p-values, can be retrieved from the online web supplement).

Keyword fingerprint of DEHP

DEHP is widely used as a plasticizer in manufacturing of a wide variety of polyvinyl chloride (PVC)-containing medical and consumer products. DEHP belongs to the class of peroxisome proliferators that induce enlargement and proliferation of peroxisomes, resulting in an elevation of fatty acid metabolism, which is characteristic of this class of chemicals [24]. This elevated level in lipid metabolism is thought to be caused by the activation of the nuclear receptor peroxisome proliferator-activated receptor (PPAR)-α, which is the key mediator of lipid metabolism [25]. Peroxisome proliferators, when chronically administered, are also known to induce tumor formation in a nongenotoxic fashion (i.e., not via direct DNA damage) [24,25], although it is not completely clear by which mechanism.

On a histopathological level, after 2 days of DEHP administration some vacuolation, hypertrophy and increases in mitotic figures were observed [20]. The observed hypertrophy is due to enlargement of liver cells caused by an

Figure 1. Keyword profiles of DMN, AB1, MPy, PBO, DEHP and VPA.



Visualization of over-represented keywords ($p \leq 0.01$), grouped in 13 biological process categories, of two genotoxic carcinogens; DMN (dotted red line) and AB1 (solid red), two nongenotoxic carcinogens; MPy (dotted blue line) and PBO (solid blue line), two peroxisome proliferators; DEHP (dotted green line) and VPA (solid green line), and of two randomly generated gene sets (150 Affymetrix probe set IDs; dotted black line and 250 Affymetrix probe set IDs; solid black line). On the y-axis the relative number of keywords, compared with the total number of keywords classified in that category, is plotted. The categorization of keywords was performed using all keywords present in at least one of the keyword fingerprints of the 11 tested compounds.
AB1: Aflatoxin B1; cat: category; DEHP: Diethylhexylphthalate; DMN: Dimethylnitrosamine; Keyw.: Keyword; MPy: Methapyrilene; PBO: Piperonylbutoxide; VPA: Valproic acid.

increase in the number of peroxisomes, most likely via activation of PPAR-α [25]. The observed vacuolation is often caused by accumulation of lipids in the cells. The profile of the keyword fingerprint of DEHP (Figure 1), shows that DEHP largely has its effect on lipid metabolism and energy metabolism, which is in agreement with the observed pathologies and with known effects of peroxisome proliferators. Keywords in the fingerprint of DEHP (Table 1 and online web supplement) support its lipogenic effect, with low p-values for keywords such as lipid metabolism, fatty acid oxidation, fatty acid β-oxidation and hypertrophy.

Keyword fingerprint of DMN

DMN is a genotoxic carcinogen (i.e., causes direct DNA damage) and is present in cigarette smoke and certain foods [26], and its observed toxicity is caused by its metabolites via methylation of DNA bases and proteins, eventually leading to the formation of tumors [19,27–29]. On a histopathological level, DMN administration results in a reactive inflammatory response, necrosis, bile duct hyperplasia and fibrosis [19,28,30].

DMN is known to induce a strong reactive inflammatory response upon administration in reaction to the onset of necrosis [19] and is followed by postnecrotic tissue remodeling [31,32]. In this process, macrophages infiltrate the liver to remove necrotic cells and activated hepatic stellate cells (HSCs) migrate to necrotic areas to secrete collagen and to promote the formation of sinusoidal wall, resulting in the formation of scar tissue [31].

The profile of the keyword fingerprint of DMN (Figure 1) is in agreement with the observation that DMN causes a strong inflammatory response following necrosis – as a relatively high number of keywords are grouped in the inflammatory and immune response category (e.g., immune system, inflammatory response and inflammation) – and that necrosis was one of the keywords with the lowest p-value (Table 1 and online web supplement). Furthermore, a relatively high number of keywords are grouped in the cell differentiation and development category (e.g., cell development,

Table 1. Keyword over-representation for diethylhexylphthalate-, dimethylnitrosamine- and methapyrilene-regulated genes .

Category	DEHP	DMN	MPy
Pathogenesis	Liver tumor, Necrosis, Cirrhosis, Hepatitis, Hypertrophy, Hepatocellular carcinoma, Cholestasis, Steatosis, Hyperplasia, Pathogenesis, Gallstone formation	Hepatocellular carcinoma, Hepatitis, Necrosis, Cirrhosis, Adenocarcinoma, Pathogenesis, Leukemia, Adenoma, Sarcoma, Fibrosis, Malignant tumor, Hyperplasia, Granuloma, Liver tumor, Angiogenesis, Hypertrophy, Atrophy, Edema, Metaplasia, Hyperplastic nodule, Ascites, Fever, Liver fibrosis, Lymphocytic leukemia, Biliary cirrhosis, Oncogenesis, Hypoplasia	Necrosis, Adenocarcinoma, Hepatocellular carcinoma, Cirrhosis, Hepatitis, Hyperplastic nodule, Leukemia, Pathogenesis, Liver tumor, Hepatocellular adenoma, Ascites, Adenoma, Sarcoma, Telangiectasia, Biliary cirrhosis, Fibrosis, Hypertrophy, Hyperplasia
Drug metabolism	Drug metabolism, Xenobiotic metabolism, Drug resistance	Drug metabolism, Drug resistance, Xenobiotic metabolism	Drug resistance, Drug metabolism, Xenobiotic metabolism, Drug transport
Oxidative stress			Response to oxidative stress, Glutathione biosynthesis, Glutathione metabolism
DNA damage and response	Mutagenesis	Mutagenesis, Cell cycle arrest, Stabilization of p53, Mismatch repair, Nucleotide excision repair, DNA damage response	DNA damage response, DNA repair, Cell cycle arrest, Stabilization of p53, Mismatch repair, Mutagenesis
Cell cycle and mitosis	Growth, Growth rate	Cell cycle, Growth, DNA synthesis, Cell proliferation, Cell growth, Growth rate, DNA replication, Cell division, DNA amplification, G2 phase, M phase	Cell cycle, Growth, Cell growth, M phase, Growth rate, G1 phase, DNA synthesis, Pachytene, G2 phase, Cell cycle checkpoint, Cell cycle, mitotic, S phase, Prophase, Transcription, mitotic, DNA replication, Cell proliferation, Mitosis, Cell cycle, Chromosome movement, Single-stranded DNA binding, Telophase, G2/M checkpoint, Chromatin remodeling, Meiosis, Division
Apoptosis and cell death		Apoptosis, Cell death, Induction of apoptosis, DNA fragmentation, Programmed cell death	Induction of apoptosis, Cell death, DNA fragmentation, Apoptosis, Programmed cell death, Caspase activation, Anti-apoptosis
Inflammation and immune response	Inflammation	Immune response, Antigen processing and presentation, Inflammatory response, Immune system, Antigen processing, Inflammation, Antigen presentation, Macrophage activation, B-cell differentiation, B-cell activation, T-cell activation, T-cell proliferation, Cytokine biosynthesis, Activation of MAPK, Chronic inflammation, Lymphocyte activation, Macrophage differentiation, Cell recognition, Mononuclear cell infiltration, Lymphocyte differentiation, Cell-mediated immune response, Humoral immune response, Cell invasion, Inflammatory cell infiltration, Monocyte activation, Antiviral response	Inflammation, T-cell proliferation, B-cell activation, Macrophage activation, Lymphocyte activation, Inflammatory response, Immune response, Antigen processing, Cytokine biosynthesis, Chronic inflammation, Antigen processing and presentation, Immune system, B-cell differentiation, Macrophage differentiation
Cell differentiation and development	Aging, Senescence	Cell development, Aging, Cell differentiation, Cell maturation, Endothelial cell differentiation, Myogenesis, Embryonic development, Cytokinesis, Brain development, Myoblast differentiation	Aging, Cell differentiation, Senescence, Epithelial cell differentiation

Table 1. Keyword over-representation for diethylhexylphthalate-, dimethylnitrosamine- and methapyrilene-regulated genes (cont.).

Category	DEHP	DMN	MPy
Steroid metabolism	Steroid metabolism, Cholesterol metabolism, Sterol metabolism	Steroid metabolism	
Lipid metabolism	Lipid metabolism, Fatty acid metabolism, Lipoprotein metabolism, Lipoprotein lipase activity, Fatty acid transport		Lipid metabolism, Lipid transport
Energy metabolism	Fatty acid oxidation, Energy metabolism, Pyruvate dehydrogenase, Fatty acid beta-oxidation, Glycolysis, Glucose metabolism, Pentose phosphate pathway, NADH metabolism, Glyoxylate metabolism, Electron transport, Tricarboxylic acid cycle	Energy metabolism, Respiratory burst, Electron transport	Energy metabolism
General cell metabolism	Metabolism, Biosynthesis, Catabolism, Transcription, Excretion, Homeostasis, Gluconeogenesis, Carbohydrate metabolism, Ethanol metabolism, Protein catabolism, Nitrogen metabolism, Protein folding, Fermentation, Glucose uptake, Amino acid metabolism, Proteoglycan metabolism, Secretion, Protein biosynthesis	Metabolism, Phagocytosis, Transcription, Biosynthesis, Translation, Secretion, Excretion, Glycosylation, Peptide transport, DNA methylation, Endocytosis, Phosphorylation, Protein stabilization, Homeostasis, Actin polymerization, Dephosphorylation, mRNA splicing, Autophosphorylation, N-glycosylation	Metabolism, Biosynthesis, Translation, Transcription, Catabolism, Ribosome, Protein biosynthesis, Heme metabolism, Response to stress, Cell homeostasis, Excretion, Protein metabolism, Phosphorylation, Histone acetylation, Dephosphorylation, Nitrogen metabolism, DNA methylation, Protein denaturation, Response to heat
Miscellaneous	Enzyme activity, Transport, Digestion, Conjugation, Blood pressure, Development, Biological process	Transport, Digestion, Cell activation, Development, Fragmentation, Cell adhesion, Cytoskeleton, Recombination, Transduction, Migration, Proteasome, Cell migration, Signal transduction, Memory, Virulence, Chemotaxis, Growth pattern, Conjugation, Cell-cell adhesion, Blood pressure, Wound healing	Fragmentation, Enzyme activity, Transport, Conjugation, Myelination, Transduction, Proteasome, Cell activation, Development, Digestion, Signal transduction, Sensitization, Cell adhesion, Heme biosynthesis, Viral replication

Results of the keyword over-representation calculation, performed with both up- and downregulated genes after DEHP, DMN and MPy administration. Note: the keywords are ranked from highest scoring keyword (i.e. with lowest p-value) to lowest scoring keyword (i.e., with highest p-value).

DEHP: Diethylhexylphthalate; DMN: Dimethylnitrosamine; MPy: Methapyrilene.

cell differentiation and cell maturation) and, together with keywords such as cell migration and cell activation (both miscellaneous category), may point to tissue remodeling, since migration, activation and differentiation of various cells occur in this process.

Interestingly, fibrosis and cirrhosis are two of the DMN-linked keywords in the pathogenesis category with low p-values. This might be correlated with the activation of HSCs by DMN, since excessive deposition of collagen can lead to fibrosis [33] and can result in cirrhosis if it becomes chronic [34].

Keyword fingerprint of MPy

MPy is a H1 histamine receptor antagonist [35] found in anti-influenza medicine that was used in the 1970s until it became known to cause cancer in rats [18,36]. MPy is a nongenotoxic carcinogen [37,38] stimulating the development of tumors, most likely via cell proliferative initiation and/or oxidative properties of MPy or its metabolites [39,40]. On a histopathological level, MPy administration results in inflammation of the liver, hepatocellular periportal necrosis, bile duct hypertrophy, periportal lipid vacuolization and apoptosis [18,36]. Interestingly, the effect of MPy appears to be species specific; no tumor formation is observed in mice, guinea-pigs, hamsters or humans [37,41,42].

It has been reported that the cytotoxicity of MPy leads to necrosis, and consequently induces regenerative cell proliferation to replace necrotic cells [18,43] and that oxidative stress is proposed to be one of the causes of the cytotoxic nature of MPy [44]. These findings reflect the profile of the keyword fingerprint of MPy (Figure 1) in that keywords related to oxidative stress, apoptosis/cell death and cell cycle/mitosis are prominent in this profile. Furthermore, keywords related to an inflammatory/immune response (e.g., inflammation and T-cell proliferation) can be correlated with the observed inflammation after MPy administration.

In short, the analysis of the keyword fingerprints of DEHP, MPy and DMN shows that the findings agree with the known effects of DEHP, MPy and DMN, and that the distinct profiles of the three compounds can be explained by the fact that they have different modes of action, leading to distinct toxic end points.

MPy: mode of action

To validate whether compound keyword fingerprints represent a realistic profile of the toxic outcome and mode of action of a particular compound, we compared in more detail the keyword fingerprint of MPy with observed toxic effects and proposed mode(s) of action of MPy, since MPy is a well-described carcinogen in the literature.

For this, we focused on the keyword over-representation results after day 7 of MPy treatment, which are shown in Table 1. At this time point, 171 genes are differentially expressed, and 118 keywords (p-value $\leq$ 0.01) are over-represented in this set of genes.

Drug metabolism

Four over-represented keywords were classified in the drug metabolism category: xenobiotic metabolism, drug resistance, drug transport and drug metabolism (Table 1). Genes associated with these keywords are P-glycoprotein/multi-drug resistance 1 *(Abcb1)*, ATP-binding cassette, subfamily C (CFTR/MRP), member 3 *(Abcc3)* and Glutathione-*S*-transferase, α-1/2 *(Gstα1/Gstα2)*, and are involved in glucoronidation and secretion of xenobiotics into bile [45–47]. These findings most likely indicate the detoxification and clearance process of MPy and its metabolites.

Inflammatory response

The presence of significant keywords that are associated with an immune reaction (e.g., inflammation and immune response; Table 1) suggests an inflammatory response upon MPy administration. Hamadeh and coworkers treated rats with 100 mg/kg/day MPy for 7 days [36], and histopathological examination of the livers of these rats showed portal inflammation and minimal mononuclear portal infiltrates, indicating the onset of an inflammatory response, which is in agreement with the keywords that we find as statistically significant in the inflammation and immune response category. Examination of genes that are associated with significant keywords in this category are involved in cell activation and proliferation, such as IL-15 and RT1 class II, locus Bb (Rt1-Bb) protein.

Pathogenesis

In the pathogenesis category, the highest scoring keyword is necrosis (Table 1 and online web supplement); this finding is in agreement with histopathological analyses of MPy experiments, carried out separately by Hamadeh [36] and Ellinger-Ziegelbauer [18]. These experiments showed the occurrence of mild necrosis after 3 days of MPy

treatment, which gradually increased in severity over time (after 7 and 14 days). Most other keywords in the pathology category are associated with carcinogenesis (e.g., hepatocellular carcinoma and liver tumor), inflammation (hepatitis) and liver degeneration (cirrhosis). These keywords are an indication of the nature of MPy as a nongenotoxic carcinogen. Nongenotoxic carcinogens are characterized by their ability to induce cell proliferation via mitogenic or cytotoxic mechanisms [43]. MPy is an example of a cytotoxic carcinogen [43], which induces cell proliferation for compensatory, regenerative cell proliferation, most likely to replace necrotic cells. The keywords that are statistically significant in the cell cycle and mitosis category (e.g., cell cycle and G1/G2/S phase) in combination with the keyword necrosis, strengthens the hypothesis that MPy induces regenerative cell proliferation. Examination of the regulated genes that are associated with keywords in the cell cycle and mitosis category shows genes involved in regulating the cell cycle: cyclin B1 *(Ccnb1)*, cyclin D1 *(Ccnd1)* and cell cycle division 2 homolog *(Cdc2a)*.

Oxidative stress

The regenerative cell proliferation abilities of MPy alone do not explain the carcinogenic nature of this compound; that is, it needs a trigger. MPy hepatotoxicity is reported to be caused by oxidative stress [44]. Oxidative stress is a cellular state in which there is an imbalance between free radicals and antioxidants. Oxidative stress can be the result of excessive formation of reactive oxygen species on the one hand, and the depletion of antioxidants, for example depletion of glutathione due to drug detoxification, on the other hand. The property of MPy causing oxidative stress might be the trigger, in combination with its regenerative cell proliferation abilities, to the onset of carcinogenesis. This can be explained by the fact that DNA replication is not error proof and, in combination with oxidative stress, increases the risk of DNA damage [43,48]. The enriched keywords response to oxidative stress and glutathione metabolism points to a reaction to prevent oxidative stress in the cell. Regulated genes associated with these keywords are heme oxygenase 1 *(Hmox1)*, metallothionein 1 *(Mt1a)* and glutathione synthetase *(Gss)*, and they are known to have a protective effect against oxidative stress [49–51].

DNA damage & response

Oxidative stress can result in a higher DNA damage incidence; this is in agreement with the observation that genes involved in DNA repair and the response to DNA damage are upregulated after MPy treatment, such as the H2A histone family, member X *(H2afx)*, and the DNA-damage inducible transcript 3 *(Ddit3)* genes. This could explain the over-represented keywords of the DNA damage and response category (e.g., DNA repair and stabilization of p53; Table 1), since MPy is a nongenotoxic carcinogen, thereby implying that DNA damage is a secondary effect, possibly caused by oxidative stress, although some studies demonstrate that MPy and/or its metabolites have a low ability to bind to DNA [52,53].

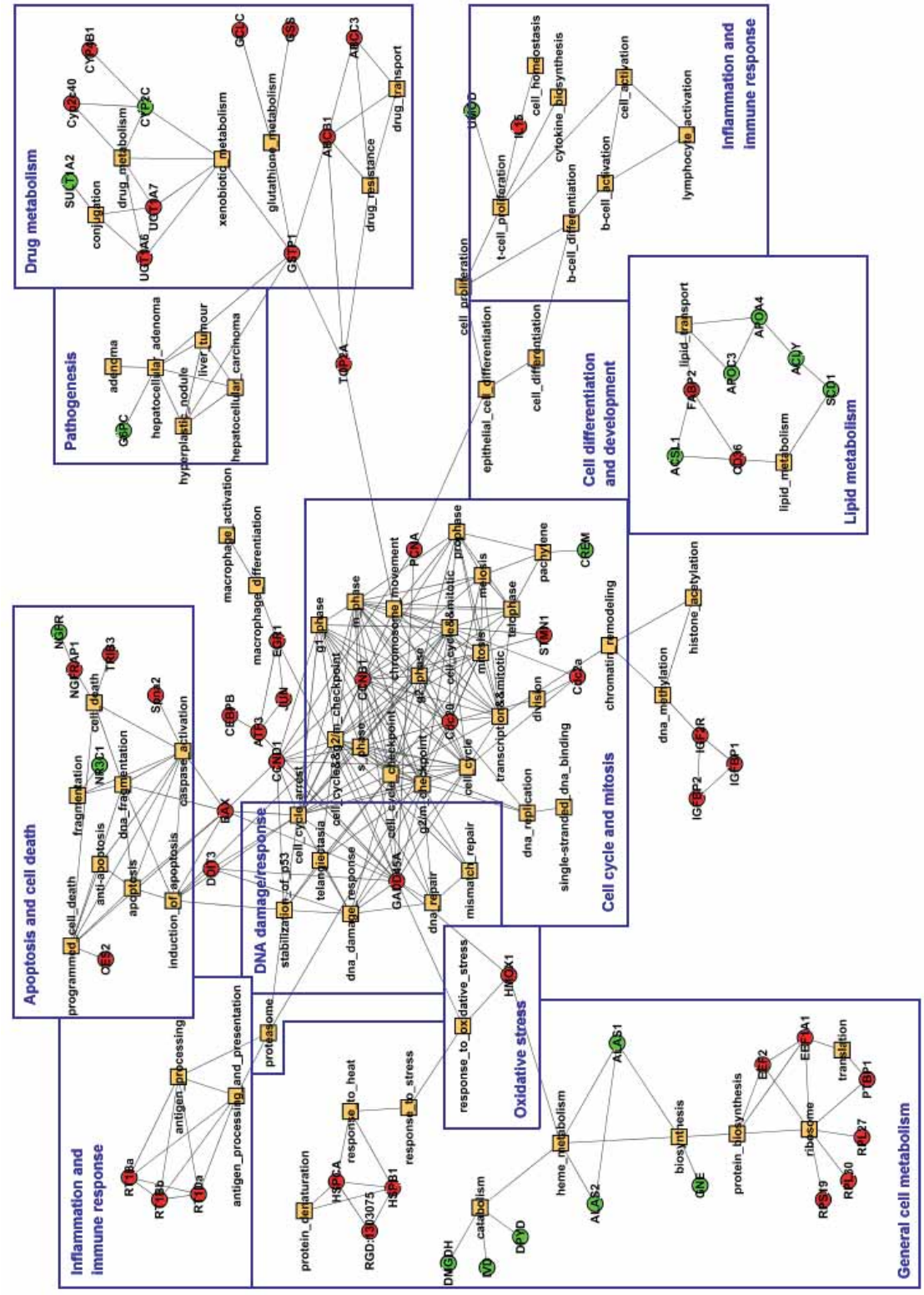
MPy or its metabolites cause mitochondrial damage, resulting in function loss, indicated by mitochondrial swelling and losses in cellular ATP. Ratra and coworkers proposed a mechanism of MPy hepatotoxicity in which mitochondrial function loss leads to a disturbed Ca^{2+} homeostasis, ultimately leading to cell death [44]. Bcl2-associated X protein *(Bax)* is one of the genes upregulated and associated with the over-represented keywords in the apoptosis and cell death category. Bax is a pro-apoptotic factor proposed to play a role in Ca^{2+} homeostasis, and thereby is a regulator of apoptosis [54].

Visualization of MPy-induced effects in a literature-based network

Applying CoPub to sets of differentially expressed genes gives insight into the separate biological processes affected by MPy, but the relationships between these processes and the most influential genes are not immediately clear. To make these relations visible, a network was generated in Cytoscape, shown in Figure 2, in which differentially expressed genes, together with over-represented keywords, are mapped. To avoid an over-complex network in which all keywords and genes have a relationship, thresholds were set to simplify the interpretation of the results. The thresholds were set to at least five co-publications between keywords and genes, with an R-scaled score of at least 45.

In the generated network, various areas representing distinct biological processes can be distinguished (Figure 2). The most influential genes and keywords are highly connected hubs. For example, the growth arrest and DNA-damage-inducible 45α *(Gadd45α)* gene is an important node between oxidative stress, DNA damage and response, apoptosis and cell death, and cell cycle and mitosis. This gene is known to have a function as a regulator of the cell cycle and is activated

Figure 2. Literature-based interaction network.



Network of differentially expressed genes and over-represented keywords after methapyrilene treatment. Red bullets: upregulated genes; green bullets: downregulated genes; squares: over-represented keywords. Network based on Medline co-publications. Threshold: at least five co-publications and R-scaled score ≥ 45.

after DNA damage, helping to trigger apoptosis [55]. The relationships between the various biological processes in the network can easily be appreciated; biologically, it makes sense that the biological processes oxidative stress and DNA damage and response are in the vicinity of each other, as well as DNA damage and response, cell cycle and mitosis, and apoptosis and cell death. Visualization of regulated genes, keywords and their relationships in a network gives an overview of the biological processes on which MPy has an effect, and is an informative tool to unravel MPy toxicity and its mechanisms of action.

Expert commentary

In this study, we introduced CoPub, a method that uses information from literature to link gene expression profiles to toxicity. With microarray data of eight hepatocarcinogens and three peroxisome proliferators [18–20], we showed that distinct keyword fingerprints of compounds can be generated using literature information.

Comparing the keyword profiles of DEHP, DMN and MPy shows that the keyword fingerprints of the three compounds agree with known observed pathologies, suggesting that the differences in the keyword fingerprints of the three compounds are based on distinct modes of action. This finding, together with the observation that the randomly generated gene sets did not result in over-represented keywords, indicates that literature information can indeed be used successfully for compound profiling and that keyword fingerprints reflect underlying biological processes.

The exact form of the keyword fingerprints that are shown in Figure 1 is partly dependent on the manual classification of the keywords in 13 high-level categories. Automatic literature-based compound profiling would benefit from structured vocabularies. Hierarchically ordered ontologies could then be used to calculate a score for every biological process, represented by a particular branch in the ontology, analogous to the analysis of gene sets for over-represented GO terms. Another possibility is to cluster the compounds on individual keywords rather than using higher level categories. We found that when we used this approach the clustering result tended to be dominated by relatively uninformative keywords, such as 'metabolism' and 'physiological process', with very low p-values (results not shown). A way around this is to use keywords that are very specific for the processes that we suspect are most important in describing toxicological events. Identification of these terms would require a well-defined set of gene expression profiles of toxic compounds with well-defined mode of action and toxicological end points.

CoPub is especially helpful for analyzing expression profiles of new compounds with an unknown mode of action. The over-represented keywords themselves already give a general overview of the affected biological processes without the need for reference compounds and an extensive knowledge by the biologist on the upregulated genes. Visualization of the keywords and the regulated genes in a network, together with the relevant abstracts in which these terms co-occur, gives an adequate impression of the affected biological processes by the tested compound and can provide leads for further experimental evaluation. For example, if the keyword profile of the tested compound shows a relative high number of keywords related to DNA damage, cell death and inflammation, then a mutagenicity test could be proposed as a follow-up study, since genotoxic carcinogens mostly hit these categories.

At this moment, our pathology thesaurus is primarily based on liver-specific pathology keywords. We could improve this thesaurus by adding other organ-specific pathology keywords to the list, and thereby broaden our perspective on compound toxicity. With this information, the assignment of organ-specific effects to compounds might be within reach.

In conclusion, compound keyword fingerprinting based on information retrieved from literature is a powerful approach for compound profiling, enabling compound toxicity evaluation and mode of action unraveling, and made more informative by visualization of over-represented keywords and differentially expressed genes in a literature network.

Future perspective

To reduce growing costs in drug development, reducing the number of compounds with undesirable characteristics as early as possible in the development process is an effective strategy. In recent years, with the development of microarray technology, toxicogenomics has emerged as an important tool for the evaluation of toxicity. The current challenge is the translation of compound-induced gene expression to the occurrence of toxicities. With the development of more sophisticated tools for data mining and gene set analysis, of which CoPub is an example, predictive toxicology and rapid compound evaluation becomes a feasible approach for reducing compound attrition in the drug development pipeline.

Acknowledgments

The authors thank Maurice Bouwhuis, Bart Heupers, Jeroen Engelberts and Pieter van Beek from SARA Computing and Networking Services (Amsterdam, The Netherlands) for updating and maintaining our CoPub database and web server [101], which was made possible by grants received from the Netherlands Bioinformatics Centre (NBIC).

Financial & competing interests disclosure

The authors have no other relevant affiliations or financial involvement with any organization or entity with a financial interest in or financial conflict with the subject matter or materials discussed in the manuscript apart from those disclosed.

No writing assistance was utilized in the production of this manuscript.

Ethical conduct of research

The authors state that they have obtained appropriate institutional review bard approval or have followed the principles outlined in the Declaration of Helsinki for all human or animal experimental investigations. In addition, for investigations involving human subjects, informed consent has been obtained from the participants involved.

Executive summary

- Toxicogenomics has emerged as a new methodology to evaluate compound toxicity and contribute to guaranteeing drug safety and reducing costs in the drug development pipeline.
- We introduced an additional approach for compound toxicity evaluation, in which literature information is used for compound profiling.
- Distinct keyword fingerprints were generated for compound profiling, representing enriched keywords in a compound-induced gene set.
- Comparing the three keyword fingerprints of diethylhexylphthalate (DEHP), methapyrilene (MPy) and dimethylnitrosamine (DMN) revealed that the findings agree with known effects of DEHP, MPy and DMN, and that the apparently distinct profiles can be explained by the fact that the three compounds have different modes of action, leading to distinct toxic end points.
- MPy-linked keywords agree with known MPy-affected biological processes and pathways, showing that keyword fingerprints represent relevant biological information.
- Visualization of MPy-linked keywords and MPy-induced genes in a literature-based network enabled us to construct a mode of toxicity proposal for MPy.
- Compound keyword fingerprinting based on information retrieved from literature is a powerful approach for compound profiling, allowing evaluation of compound toxicity and analysis of the mode of action.

Bibliography

Papers of special note have been highlighted as either of interest (•) or of considerable interest (••) to readers.

1. Walmsley RM: Genotoxicity screening: the slow march to the future. *Expert Opin. Drug Metab. Toxicol.* 1(2), 261–268 (2005).
2. Afshari CA, Nuwaysir EF, Barrett JC: Application of complementary DNA microarray technology to carcinogen identification, toxicology, and drug safety evaluation. *Cancer Res.* 59(19), 4759–4760 (1999).
3. Nuwaysir EF, Bittner M, Trent J, Barrett JC, Afshari CA: Microarrays and toxicology: the advent of toxicogenomics. *Mol. Carcinog.* 24(3), 153–159 (1999).
4. Castle AL, Carver MP, Mendrick DL: Toxicogenomics: a new revolution in drug safety. *Drug Discov. Today* 7(13), 728–736 (2002).
5. Subramanian A, Tamayo P, Mootha VK *et al.*: Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl Acad. Sci. USA* 102(43), 15545–15550 (2005).
6. Yi M, Horton JD, Cohen JC, Hobbs HH, Stephens RM: WholePathwayScope: a comprehensive pathway-based analysis tool for high-throughput data. *BMC Bioinformatics* 7, 30 (2006).
7. Currie RA, Bombail V, Oliver JD *et al.*: Gene ontology mapping as an unbiased method for identifying molecular pathways and processes affected by toxicant exposure: application to acute effects caused by the rodent non-genotoxic carcinogen diethylhexylphthalate. *Toxicol Sci.* 86(2), 453–469 (2005).
8. Jenssen TK, Laegreid A, Komorowski J, Hovig E: A literature network of human genes for high-throughput analysis of gene expression. *Nat. Genet.* 28(1), 21–28 (2001).
9. Jelier R, Jenster G, Dorssers LC *et al.*: Text-derived concept profiles support assessment of DNA microarray data for acute myeloid leukemia and for androgen receptor stimulation. *BMC Bioinformatics* 8, 14 (2007).
10. Chaussabel D, Sher A: Mining microarray expression data by literature profiling. *Genome Biol.* 3(10), research0055.1–research0055.16 (2002).
11. Chagoyen M, Carmona-Saez P, Shatkay H, Carazo JM, Pascual-Montano A: Discovering semantic features in the literature: a foundation for building functional associations. *BMC Bioinformatics* 7, 41 (2006).
12. Kuffner R, Fundel K, Zimmer R: Expert knowledge without the expert: integrated analysis of gene expression and literature to

derive active functional contexts. *Bioinformatics* 21(Suppl. 2), ii259–ii267 (2005).

13. Glenisson P, Antal P, Mathys J, Moreau Y, De Moor B: Evaluation of the vector space representation in text-based gene clustering. *Pac. Symp. Biocomput.* 391–402 (2003).
14. Rubinstein R, Simon I: MILANO – custom annotation of microarray results using automatic literature searches. *BMC Bioinformatics* 6, 12 (2005).
15. Blaschke C, Oliveros JC, Valencia A: Mining functional information associated with expression arrays. *Funct. Integr. Genomics* 1(4), 256–268 (2001).
16. Alako BT, Veldhoven A, van Baal S *et al.*: CoPub Mapper: mining MEDLINE based on search term co-publication. *BMC Bioinformatics* 6, 51 (2005).
17. Ding J, Berleant D, Nettleton D, Wurtele E: Mining MEDLINE: abstracts, sentences, or phrases? *Pac. Symp. Biocomput.* 7, 326–337 (2002).
18. Ellinger-Ziegelbauer H, Stuart B, Wahle B, Bomann W, Ahr HJ: Comparison of the expression profiles induced by genotoxic and nongenotoxic carcinogens in rat liver. *Mutat. Res.* 575(1–2), 61–84 (2005).
    •• Expression profiling of genotoxic and nongenotoxic carcinogens to identify characteristic sets of genes that represent defined biological pathways to predict a genotoxic or nongenotoxic carcinogenic potential of a compound.
19. Ellinger-Ziegelbauer H, Stuart B, Wahle B, Bomann W, Ahr HJ: Characteristic expression profiles induced by genotoxic carcinogens in rat liver. *Toxicol. Sci.* 77(1), 19–34 (2004).
    • Expression profiling of genotoxic carcinogens to identify genes and pathways that are indicative of early events in tumorigenesis.
20. Jolly RA, Goldstein KM, Wei T *et al.*: Pooling samples within microarray studies: a comparative analysis of rat liver transcription response to prototypical toxicants. *Physiol. Genomics* 22(3), 346–355 (2005).
    • Discusses the expression profiles of three well-characterized peroxisome proliferators in rat liver.
21. Maglott D, Ostell J, Pruitt KD, Tatusova T: Entrez Gene: gene-centered information at NCBI. *Nucleic Acids Res.* 35(Database issue), D26–D31 (2007).
22. Parkinson H, Kapushesky M, Shojatalab M *et al.*: ArrayExpress – a public database of microarray experiments and gene expression profiles. *Nucleic Acids Res.* 35(Database issue), D747–D750 (2007).
23. Shannon P, Markiel A, Ozier O *et al.*: Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13(11), 2498–2504 (2003).
24. Rusyn I, Peters JM, Cunningham ML: Modes of action and species-specific effects of di-(2-ethylhexyl)phthalate in the liver. *Crit. Rev. Toxicol.* 36(5), 459–479 (2006).
25. Gonzalez FJ, Peters JM, Cattley RC: Mechanism of action of the nongenotoxic peroxisome proliferators: role of the peroxisome proliferator-activator receptor α. *J. Natl Cancer Inst.* 90(22), 1702–1709 (1998).
26. Brunnemann KD, Fink W, Moser F: Analysis of volatile *N*-nitrosamines in mainstream and sidestream smoke from cigarettes by GLC-TEA. *Oncology* 37(4), 217–222 (1980).
27. Barrows LR: Methylation of DNA guanine via the 1-carbon pool in dimethylnitrosamine-treated rats. *Mutat. Res.* 173(1), 73–79 (1986).
28. George J, Rao KR, Stern R, Chandrakasan G: Dimethylnitrosamine-induced liver injury in rats: the early deposition of collagen. *Toxicology* 156(2–3), 129–138 (2001).
29. Gombar CT, Zubroff J, Strahan GD, Magee PN: Measurement of 7-methylguanine as an estimate of the amount of dimethylnitrosamine formed following administration of aminopyrine and nitrite to rats. *Cancer Res.* 43(11), 5077–5080 (1983).
30. Schook LB, Lockwood JF, Yang SD, Myers MJ: Dimethylnitrosamine (DMN)-induced IL-1 beta, TNF-α, and IL-6 inflammatory cytokine expression. *Toxicol. Appl. Pharmacol.* 116(1), 110–116 (1992).
31. Jin YL, Enzan H, Kuroda N *et al.*: Tissue remodeling following submassive hemorrhagic necrosis in rat livers induced by an intraperitoneal injection of dimethylnitrosamine. *Virchows Arch.* 442(1), 39–47 (2003).
32. Jin YL, Enzan H, Kuroda N *et al.*: Vascularization in tissue remodeling after rat hepatic necrosis induced by dimethylnitrosamine. *Med. Mol. Morphol.* 39(1), 33–43 (2006).
33. Sato M, Suzuki S, Senoo H: Hepatic stellate cells: unique characteristics in cell biology and phenotype. *Cell Struct. Funct.* 28(2), 105–112 (2003).
34. Guyot C, Lepreux S, Combe C *et al.*: Hepatic fibrosis and cirrhosis: the (myo)fibroblastic cell subpopulations involved. *Int. J. Biochem. Cell Biol.* 38(2), 135–151 (2006).
35. Noguchi S, Inukai T, Kuno T, Tanaka C: The suppression of olfactory bulbectomy-induced muricide by antidepressants and antihistamines via histamine H1 receptor blocking. *Physiol. Behav.* 51(6), 1123–1127 (1992).
36. Hamadeh HK, Knight BL, Haugen AC *et al.*: Methapyrilene toxicity: anchorage of pathologic observations to gene expression alterations. *Toxicol. Pathol.* 30(4), 470–482 (2002).
37. Mirsalis JC: Genotoxicity, toxicity, and carcinogenicity of the antihistamine methapyrilene. *Mutat. Res.* 185(3), 309–317 (1987).
38. Steinmetz KL, Tyson CK, Meierhenry EF, Spalding JW, Mirsalis JC: Examination of genotoxicity, toxicity and morphologic alterations in hepatocytes following *in vivo* or *in vitro* exposure to methapyrilene. *Carcinogenesis* 9(6), 959–963 (1988).
39. Casciano DA, Talaska G, Clive D: The potent hepatocarcinogen methapyrilene induces mutations in L5178Y mouse lymphoma cells in the apparent absence of DNA adduct formation. *Mutat. Res.* 263(2), 127–132 (1991).
40. Ratra GS, Powell CJ, Park BK, Maggs JL, Cottrell S: Methapyrilene hepatotoxicity is associated with increased hepatic glutathione, the formation of glucuronide conjugates, and enterohepatic recirculation. *Chem. Biol. Interact.* 129(3), 279–295 (2000).
41. Brennan LM, Creasia DA: The effects of methapyrilene hydrochloride on hepatocarcinogenicity and pentobarbital-induced sleeping time in rats and mice. *Toxicol. Appl. Pharmacol.* 66(2), 252–258 (1982).
42. Lijinsky W, Knutsen G, Reuber MD: Failure of methapyrilene to induce tumors in hamsters or guinea pigs. *J. Toxicol. Environ. Health* 12(4–6), 653–657 (1983).
43. Cunningham ML: Role of increased DNA replication in the carcinogenic risk of nonmutagenic chemical carcinogens. *Mutat. Res.* 365(1–3), 59–69 (1996).
44. Ratra GS, Morgan WA, Mullervy J, Powell CJ, Wright MC: Methapyrilene hepatotoxicity is associated with oxidative stress, mitochondrial disfunction and is prevented by the $Ca2^+$ channel blocker verapamil. *Toxicology* 130(2–3), 79–93 (1998).
45. Pastore A, Federici G, Bertini E, Piemonte F: Analysis of glutathione:

implication in redox and detoxification. *Clin. Chim. Acta* 333(1), 19–39 (2003).
46. Rost D, Konig J, Weiss G *et al.*: Expression and localization of the multidrug resistance proteins MRP2 and MRP3 in human gallbladder epithelia. *Gastroenterology* 121(5), 1203–1208 (2001).
47. Schinkel AH: The physiological function of drug-transporting P-glycoproteins. *Semin. Cancer Biol.* 8(3), 161–170 (1997).
48. Cohen SM: Role of cell proliferation in regenerative and neoplastic disease. *Toxicol. Lett.* 82–83, 15–21 (1995).
49. Anderson ME: Glutathione: an overview of biosynthesis and modulation. *Chem. Biol. Interact.* 111–112, 1–14 (1998).
50. Kang YJ: Metallothionein redox cycle and function. *Exp. Biol. Med. (Maywood)* 231(9), 1459–1467 (2006).
51. Peng L, Mundada L, Stomel JM *et al.*: Induction of heme oxygenase-1 expression inhibits platelet-dependent thrombosis. *Antioxid. Redox Signal.* 6(4), 729–735 (2004).
52. Lampe MA, Kammerer RC: Cytochrome P-450 dependent binding of methapyrilene to DNA *in vitro*. *Carcinogenesis* 8(10), 1525–1529 (1987).
53. Lampe MA, Kammerer RC: Species differences in the metabolism and macromolecular binding of methapyrilene: a comparison of rat, mouse and hamster. *Xenobiotica* 20(12), 1269–1280 (1990).
54. Carvalho AC, Sharpe J, Rosenstock TR *et al.*: Bax affects intracellular $Ca2^{+}$ stores and induces $Ca2^{+}$ wave propagation. *Cell Death Differ.* 11(12), 1265–1276 (2004).
55. Gramantieri L, Chieco P, Giovannini C *et al.*: GADD45-α expression in cirrhosis and hepatocellular carcinoma: relationship with DNA repair and proliferation. *Hum. Pathol.* 36(11), 1154–1162 (2005).

Websites

101. CoPub
http://services.nbic.nl/cgi-bin/copub/CoPub.pl
102. National Center for Biotechnology Information (NCBI) Entrez Gene
www.ncbi.nih.gov/entrez/query.fcgi?db=gene
103. The Gene Ontology (GO) project
www.geneontology.org
104. KEGG: Kyoto Encyclopedia of Genes and Genomes
www.genome.jp/kegg
105. Encyclopedia of Human Genes and Metabolism
http://humancyc.org
106. Reactome
www.genomeknowledge.org
107. Medline/PubMed Baseline Distribution
www.nlm.nih.gov/bsd/licensee/2007_stats/baseline_doc.html
108. ArrayExpress
www.ebi.ac.uk/arrayexpress
109. Affymetrix
www.affymetrix.com
110. The R Project for Statistical Computing
www.r-project.org
111. Cytoscape: Analyzing and Visualizing Network Data
www.cytoscape.org