



Funded by the Seventh Framework
Programme of the European Union



Project full title:
Hepatic and Cardiac Toxicity Systems modelling

Project acronym:
HeCaTos

Collaborative project
HEALTH.2013.1.3.-1:
Modelling toxic response in case studies for predictive human safety assessment

FP7-HEALTH-2013-INNOVATION-1-602156-HeCaTos

Deliverable Report D2.1:
Report on initial molecular networks for liver/heart toxicity from literature data

Edited by
Ralf Herwig, Johannes Schuchhardt, Arif Malik
Work package 2

Due date of deliverable: M12
Actual submission date: November 2014

Start date of project: October, 2013

Duration: 60 months

Maastricht University (UM)

Project co-funded by the European Commission within the 7th Framework Programme (2013-2018)		
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)	

Contributions to deliverable - Internal review procedure

Deliverable produced by:	Date:
Johannes Schuchhardt - Partner MD	04.11.2104
Ralf Herwig - Partner MPIMG	05.11.2014
Arif Malik - Partner MD	06.11.2014
Arif Malik - Partner MD	07.11.2014
Deliverable internally reviewed by:	Date:
Jos Kleinjans - Partner UM	10.11.2014

Contents

Publishable Summary.....	3
Objectives	3
Introduction	4
1. Results	5
1.1. Ranking of toxic compounds for construction of initial networks of toxicity	5
1.2. Literature based construction of gene and protein networks associated with toxic compounds	5
1.3. Identifying KEGG Pathways for acetaminophen	6
1.4. Microarray analysis	7
1.5. Interaction network construction - ConsensusPathDB, CAPPIC and IntScore	10
1.6. Network analysis	12
1.7. Toxicity modules	13
Difficulties	17
References	18
Annex: Top-50 scoring genes per compound derived from A) literature mining and B) gene expression analysis	19

PUBLISHABLE SUMMARY

In a proof-of principle approach we investigate the toxic key compounds acetaminophen, rifampin, azathioprine (liver toxins) and doxorubicin HCL, cyclophosphamide (heart toxins) with the goal to construct networks of molecular species that are predictive for liver and heart toxicity.

Employing methods of literature mining in combination with statistical investigation of large scale OMICS data sets we construct core gene lists associated with heart and liver toxicity. Using graph based methods for the construction of gene and protein networks we obtain key hubs of toxic action.

OBJECTIVES

Aim of the work performed in year one is to construct networks of molecular species that are predictive for liver and heart toxicity in order to use these networks in WPs 11 and 12 for integrative data analysis and in WP2 for the basis of molecular models. This goal was achieved in cooperation between partner 10 (Partner MPIMG) and 13 (Partner MD).

As a first proof-of-concept, WP2 partners investigated five compounds of primary interest for the consortium: acetaminophen, rifampin, azathioprine (liver toxins) and doxorubicin HCL, cyclophosphamide (heart toxins). Criteria for selection were severity of toxicity, availability of omics data and relevance for the consortium.

The approach to derive the compound-specific predictive networks for liver/heart toxicity consists of the following steps:

1. For each compound derive a weighted prioritized list of related genes/proteins through literature mining (Partner MD);
2. For each compound derive a weighted prioritized list of related genes/proteins through microarray data analysis (Partner MPG);
3. Derive a high-quality protein-protein interaction (PPI) network from ConsensusPathDB (Partner MPIMG);
4. Combine the weights from literature mining and gene expression analysis and weight the nodes in the PPI network accordingly (Partner MPIMG);
5. For each compound compute a sub network that agglomerates high-scoring nodes of the network (Partner MPIMG).

INTRODUCTION

MicroDiscovery has set up a processing pipeline for literature mining employing refined statistical methods for the context dependent identification of biomolecules and their interaction. MPIMG has set up a comprehensive resource for identifying relevant genes and proteins and related pathway information from large OMICS studies. In our work we combine these methods in order to identify networks of molecular species predicting liver and heart toxicity.

The overall approach is illustrated in Figure 1. Starting from a pre-defined list of toxic compounds (the selection will be described below) parallel searches are performed for identifying associated biomolecules. One strategy for identifying biomolecules is based on literature mining, the second strategy is based on OMICS data analysis. In a further step this core set of data is extended by information on protein-protein interactions integrated from ConsensusPathDB. The resulting network is investigated for hubs using graph algorithms. Details on how the different steps are performed are presented below.

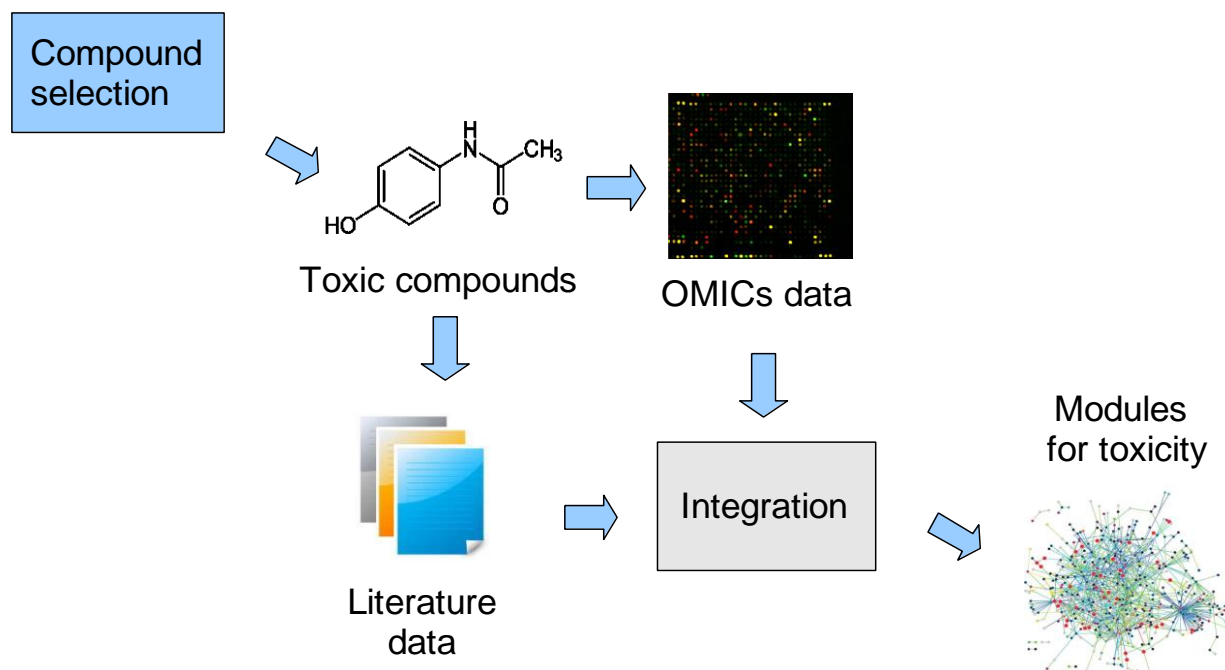


Figure 1. General workflow for identifying modules of toxicity. Following the initial process of compound selection, parallel branches of information retrieval are followed. Literature based identification of related genes and proteins and OMICS data based identification. These core components are extended based on protein-protein interaction information. The resulting network is investigated by graph algorithms for the identification of densely connected sub-networks (hubs or cliques). These sub-networks are the identified modules of toxicity and form the basis for further research in WP2.

1. RESULTS

Literature mining

1.1. Ranking of toxic compounds for construction of initial networks of toxicity

In a first step a prioritization of toxic compounds for network construction was achieved. A set of criteria for compound ranking was formulated during the project board meetings.

The pre-selected list as generated by the board was subjected to the following ranking criteria:

- Availability of OMICS data for analysis;
- Availability of PK-PD models for estimating drug dosage to be provided by WP4 (RWTH);
- Availability of specific literature on toxicity effects on liver or heart respectively.

Table 1: Combining information from different sources the top ranking compounds for liver toxicity were identified

	TG GATES Liver Human	ToxDB (liver)	CTD Interac. Genes	Molecular Models WP4	Lit Liver/ Heart Ratio	Lit Liver All Ratio(%)	Rank
Acetaminophen	24	yes	9205	yes	12,612	10,795	1
Rifampin	24	yes	188	yes	16,638	3,347	2
Azathioprine	24	yes	75	yes	2,181	2,697	3
Isoniazid	24	yes	59	no	44,952	4,769	>3
Valproic acid	24	yes	11252	no	8,846	3,464	>3
Diclofenac	24	yes	510	no	4,571	3,048	>3
Phenytoin	24	yes	404	no	2,600	2,061	>3
Erythromycin	12 (Ethysucc.)	yes	179	soon	3,198	1,014	>3
Pravastatin	no	no	58	yes	1,092	1,981	>3
Methotrexate	no	kidney	1349	no	3,204	2,483	>3
Piroxicam	no	kidney	555	no	3,188	1,492	>3

A similar approach was taken for ranking heart toxicity compounds. For the generation of initial networks we identified as top three compounds for liver:

- Acetaminophen;
- Azathioprine;
- Rifampin.

The top two compounds for heart are:

- Doxorubicin;
- Cyclophosphamide.

1.2. Literature based construction of gene and protein networks associated with toxic compounds

Starting from the top five compounds selected for liver and heart toxicity we performed a differential investigation of gene and protein entities significantly associated with the compound in the context of liver or heart toxicity. This leads to compound specific lists of associated biomolecules as exemplarily given in Table 2 more complete listing giving the top fifty genes and proteins for each of the five compounds investigated is reported in the appendix

Table 2: Short listing of the most significant genes/proteins associated with rifampin as identified based on differential literature mining. The last three columns give some statistical information on the search hit. Genes/proteins are ranked according to their p-value of association with the compound rifampin.

Symbol	Description	Syns	NTest	NRef	pv
CYP3A4	cytochrome P450, family 3, subfamily A	CYP3A4(94);CYP3A(24);cyt	122	7081	0
NR1I2	nuclear receptor subfamily 1, group I, n	PXR(98);SXR(9);NR1I2(3);p	115	4034	0
AADAC	arylacetamide deacetylase	AADAC(14);arylacetamide	15	67	1,9E-102
SLCO1B1	solute carrier organic anion transporter	OATP1B1(12);OATP-C(3);SL	21	1468	4,43E-89
NR1I3	nuclear receptor subfamily 1, group I, n	CAR(17);NR1I3(1)	18	1317	1,31E-76
SLC10A1	solute carrier family 10 (sodium/bile ac	NTCP(11);SLC10A1(2)	13	532	3,07E-64
ABCB11	ATP-binding cassette, sub-family B (MD	BSEP(11);ABCB11(1);ABCB	15	1339	2,9E-62
SLCO1B3	solute carrier organic anion transporter	OATP8(6);OATP1B3(4);orga	12	718	1,24E-55
GPT	glutamic-pyruvate transaminase (alanin	ALT(20);alanine aminotran	24	15685	4,21E-53
G6PC	glucose-6-phosphatase, catalytic subun	G6Pase(13);glucose 6-phos	15	2808	4,38E-52
CYP2B6	cytochrome P450, family 2, subfamily B	CYP2B6(8);P450(7);CYP2B(16	4783	2,44E-48
MIR34A	microRNA 34a	miR-34a(11)	11	1162	1E-45
CYP2E1	cytochrome P450, family 2, subfamily E	CYP2E1(11);cytochrome P-	12	2692	7,09E-41
CYP1A2	cytochrome P450, family 1, subfamily A	CYP1A2(12)	12	3190	5,49E-39

In order to select more specific genes, we compare the results with the reference set using a Fisher enrichment statistic. The 14 most specific genes for rifampin and liver toxicity are shown in Table 2 (for a more complete listing see Annex). For each gene we calculate a p-value using Fisher's test statistic. The p-value is a reliable measure of the strength and specificity of the association of the gene or protein with rifampin in the context of liver toxicity.

Among the top hits for rifampin we see different members from CYP protein-family. CYPs are the major enzymes involved in drug metabolism. Among others, they are responsible for drug deactivation or they facilitate excretion from the body.

Similar encouraging results are found for acetaminophen (see Annex): prominent among the top ranked proteins are several members of the cytochrome P450 family. Also among the top hits are UDP-glucurotyl transferase and sulfo transferase that typically metabolize acetaminophen leading to secretion in the urine. The list of compounds identified is very reasonable and forms a good starting point for the subsequent steps of network construction.

1.3. Identifying KEGG Pathways for acetaminophen

With the most significant genes (genes with p-values < 1E-05) from literature mining we performed a pathway enrichment analysis. The most significant KEGG pathways are: Drug metabolism, Metabolism of xenobiotics by cytochrome P450 and Chemical carcinogenesis (with p-value 1E-15). The top KEGG pathway is 'Drug metabolism'. Especially the metabolism of Codeine, Methadone, Lidocaine and Felbamate seem to involve similar mechanisms as hepatotoxicity of acetaminophen.

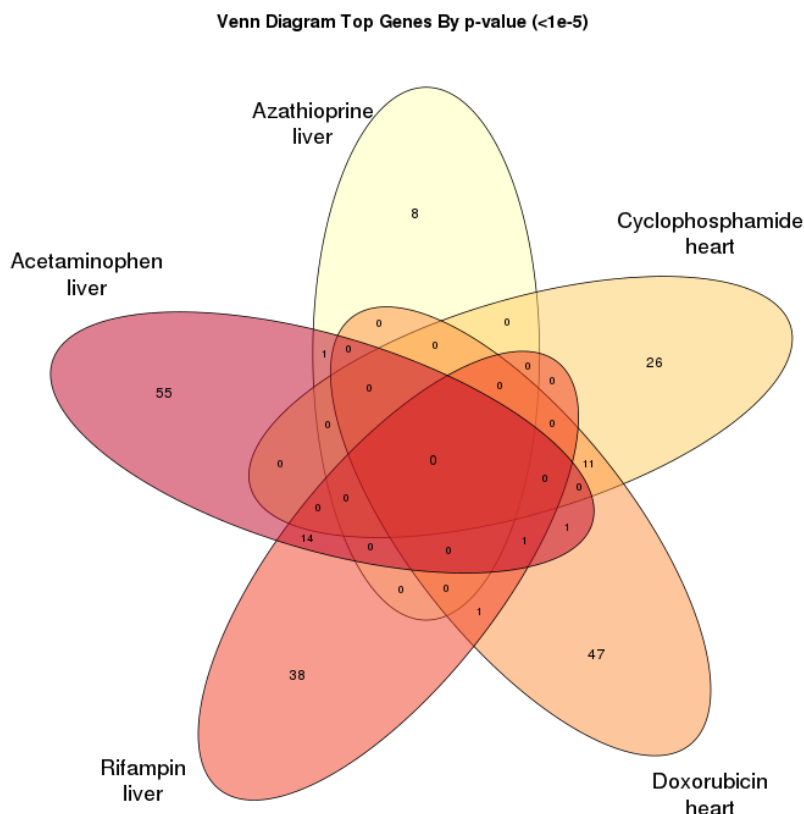


Figure 2: Investigating commonalities and differences among the gene and protein sets identified for the five top compounds shows there is no overlap between all five compounds.

Text mining was performed for all five compounds. The results are provided as Excel tables. There is a small overlap of the two heart related compounds doxorubicin and cyclophosphamide as well as an overlap between acetaminophen and rifampin. Azathioprine has shown almost no overlap to any of the other compounds (Figure 2).

1.4. Microarray analysis

MPI-MG selected appropriate data sets for microarray analysis. While it is in principle preferable to work with human data, for the cardio-toxic compounds there is no human data available currently and thus we had to work with rat data. With all compounds we chose the strategy of analyzing data derived from experiments with high compound concentrations and long exposure times in order to capture as much effects as possible. Table 3 lists the experiments that were used for analysis.

Compound	CAS number	Exposure time	Dosage	Organism	Target organ	Study
Acetaminophen	103-90-2	1 day	5000 μ M	Human	hepatocytes	TG-GATES
Rifampin	13292-46-1	1 day	70 μ M	Human	hepatocytes	TG-GATES
Azathioprine	446-86-6	1 day	72.8 μ M	Human	hepatocytes	TG-GATES
Doxorubicin HCL	23214-92-8	5 days	3 mg/kg	Rat	heart	DrugMatrix
Cyclophosphamide	50-18-0	5 days	25 mg/kg	Rat	heart	DrugMatrix

Table 3: Microarray experiments used for the study.

Raw gene expression data was derived from the diXa platform that builds also the basis for the HeCaToS data warehouse (WP9). Differential expression analysis was carried according to published protocols of MPIMG. Analysis steps include:

- Remapping of Affymetrix oligoprobes to the latest genome build;
- QC and pre-processing using the GCRMA method;
- Statistical testing using Student's t-test;
- Correction for multiple testing using FDR.

In all cases we compared replicated treatment experiments with replicated control experiments and derived a fold-change and a p-value indicating the significance of the fold-change as judged by Student's t-test. The mostly differentially expressed genes for the five compounds are listed in the appendix, complete results have been uploaded to the HeCaToS web repository.

Example: microarray results for rifampin

As an example we present the results of the microarray analysis for rifampin. Applying, for example rather strict filtering with $q\text{-value} < 0.2$ and $\text{fold-change} \geq 2$, yields 28 genes with highest significance for dysregulation (Table 4).

Table 4: Highly differentially expressed genes upon rifampin treatment.

Gene Symbol	Fold-change (\log_2)	P-value	Q-value
CYP3A4	7,875	0,000117584	0,124
CYP3A7	6,215	0,000228469	0,162
THRSP	4,236	0,001	0,178
TSKU	3,151	0,001	0,178
C19orf80	2,083	0,003	0,199
CYP3A5	1,97	0,002	0,199
ALAS1	1,883	0,001	0,183
CYP2C8	1,785	0,001	0,199
POR	1,778	0,004	0,199
SLC27A3	-1,764	0,001	0,199
CYP3A43	1,72	0,003	0,199
DIO1	1,595	0,002	0,199
RP11-372E1.4	1,513	0,001	0,199
CYP2C9	1,419	1,59E-05	0,124
VLDLR	1,346	0,001	0,199
MBL2	1,285	0,004	0,199

SERPINB9	1,271	1,43E-05	0,124
CXCL2	-1,262	0,000283348	0,169
INSIG1	1,24	0,002	0,199
RHOF	1,189	0,002	0,199
SEC14L4	1,186	0,001	0,199
IGF1	-1,173	0,003	0,199
UPP1	1,135	0,002	0,199
RRAD	-1,129	0,002	0,199
PRODH2	1,097	0,002	0,199
CD14	1,091	0,003	0,199
CYP2C19	1,078	0,002	0,199
SLC7A5	1,007	0,001	0,199

As can be seen from the short list CYP3A4 is the top-scoring gene. This is well documented in the literature. CYP3A4 is induced by a wide variety of ligands. These ligands (among which is rifampicin) bind to the pregnane X receptor (PXR). The activated PXR complex forms a heterodimer with the retinoid X-receptor (RXR), which binds to the CYP3A4 gene. Further CYP family genes that are activated by rifampin with the microarray experiment and that are reported in the literature are CYP3A7 and CYP3A5 (Yamashita et al., 2013).

Thus, the computed p-value seems a reasonable measure for dysregulation of the respective gene and is kept for node scoring for the subsequent network analysis.

Figure 3 illustrates the overlap of the gene lists derived by literature mining and gene expression analysis (point 1 above). There is a considerable overlap between gene lists derived from literature mining and microarray analysis for the hepatotoxic compounds (~30%-40%) and a lesser overlap for the cardiotoxic compounds (~10%-15%). This reflects the fact that the cardiotoxic compounds were evaluated with DrugMatrix data, i.e. gene expression data from rat, while literature mining searched for associations in the human context.

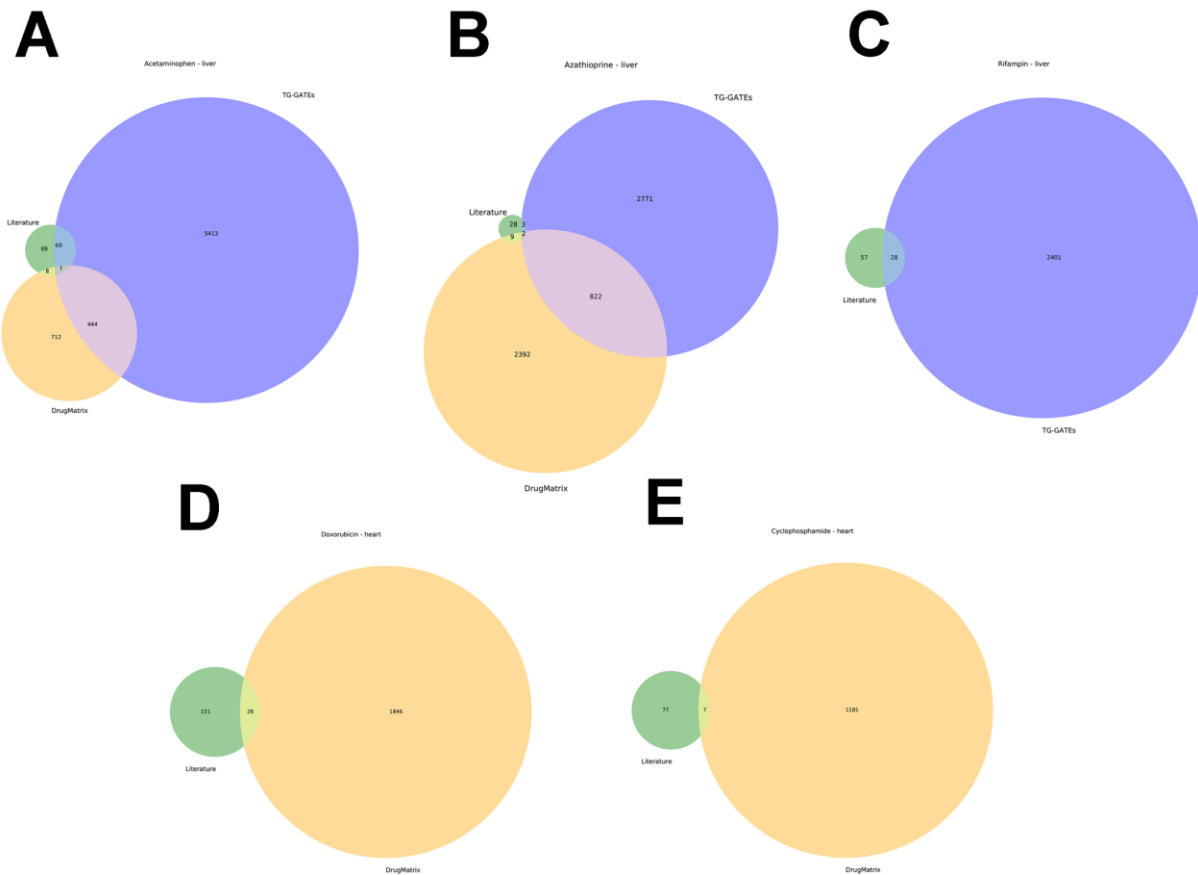


Figure 3: Overlap of significant genes (p-value<0.05) derived from literature mining (green circles), TG-GATES (blue circles) and DrugMatrix (orange circles) microarray data sets. A: acetaminophen, B: azathioprine, C: Rifampin, D: doxorubicin, E: cyclophosphamide.

1.5. Interaction network construction - ConsensusPathDB, CAPPIC and IntScore

After computation of p-values from literature mining and gene expression which yields a priority scoring for the human genes, we need to derive a high quality network that provides interactions of the genes. The interaction network used in WP2 was retrieved from MPIMG's interaction resource ConsensusPathDB (<http://consensuspathdb.org>). Since it is known that protein-protein interaction (PPI) networks contain a large part of false positive interactions, the quality assessment of individual PPIs is particularly critical for the ConsensusPathDB because this meta-database integrates PPIs from 19 different databases comprising a large network of 221,328 binary PPIs.

There are several methods available that judge quality of PPIs, for example:

- Unsupervised methods, i.e. topological methods that try to extract a confidence value for a given PPI by exploring local network topology, and
- Supervised methods, i.e. annotation-based methods that judge confidence of PPIs by exploring joint occurrences of the proteins in published sources, e.g. GO trees, literature, pathways.

MPIMG has developed a topological method for assessing confidence of PPIs, called CAPPIC (cluster-based assessment of protein-protein interaction confidence) which is based on Markov Clustering of the interaction network and subsequent local assessment of interactions according to the cluster structure (Kamburov et al., 2012). Furthermore, this and six other measures have been implemented into a web interface, called IntScore, that computes the individual confidence score as well as an aggregated score

for each PPI of a given network (Kamburov et al., 2012). These tools have been developed prior to the HeCaToS project and were used for analysis.

We have computed the aggregated confidence scores by applying IntScore to the original PPI network of ConsensusPathDB and retrieved a high-quality network consisting of 9,533 nodes with 80,422 interactions (Figure 4A). The node degree in the remaining network follows a power law distribution which is postulated for biological networks, i.e. a network that has few hubs with many connections and many nodes with rather few connections (Figure 4B). Additionally, the IntScore functionality was incorporated into ConsensusPathDB: The values resulting from the aggregated confidence scoring are displayed for each interaction so that the user can decide whether or not to work further on a particular PPI (Figure 4C).

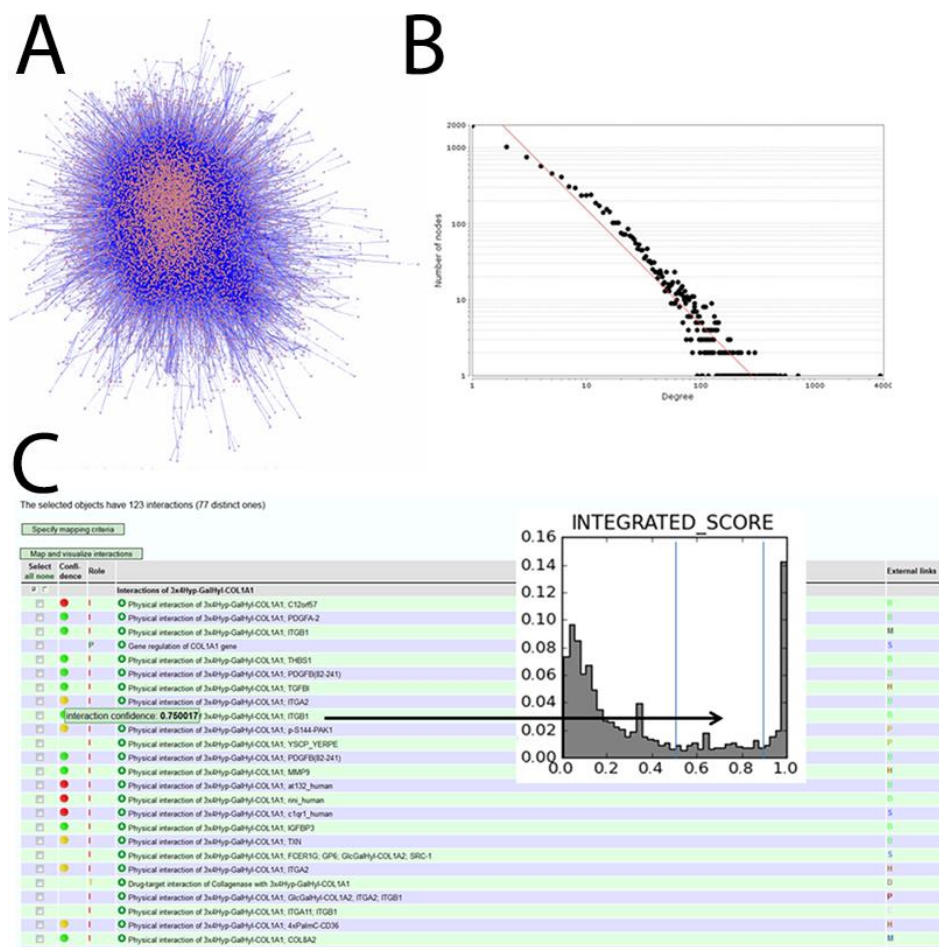


Figure 4: PPI network A) High-quality PPI network with 9,533 nodes and 80,422 interactions derived from ConsensusPathDB after confidence scoring using IntScore. B) Node degree distribution of the human protein-protein interaction network derived from high quality interactions of the ConsensusPathDB. The node distribution shows the typical power law distribution $y = a \cdot x^b$ with parameters $a = 4896.5$ and $b = -1.504$. C) Interaction confidence scores are integrated in the ConsensusPathDB as a color-coded aid for the user.

1.6. Network analysis

Given the PPI network (point 3 above) and, for each gene/protein, the p-values derived from literature mining (point 1) and microarray analysis (point 2), network analysis can be applied in order to compute the predictive toxicity networks modules from the large PPI network. As an appropriate algorithm to compute such network modules partners have identified the BioNet approach.

BioNet is available as an R-package which could easily be integrated in MPG's computational pipeline. The algorithm needs as input a weighted PPI network. Weights for the corresponding genes/proteins were derived by combining the two p-values described above and overlaid to the PPI network using the following p-value transformation: Given the summarized p-values derived from literature mining and gene expression measurements, these p-values are transformed into a mixture of two Beta-functions in order to separate signal ($B(a,1)$ distribution) from noise ($B(1,1)$ uniform distribution). Such a function has the form:

$$f(x/a, \lambda) = \lambda + (1-\lambda) ax^{a-1} \text{ with } 0 < x \leq 1; 0 < a < 1.$$

and the function uses the initial p-values to derive estimates for a and λ .

In the next step, given a certain false-discovery rate, $t(\text{FDR})$, this fit is used to derive for each node x a node score, $S(x)$, by:

$$S(x) = (a-1) (\log(x) - \log(t(\text{FDR}))).$$

Once the nodes of the PPI network have been weighted according to these values the BioNet approach applies linear integer programming in order to derive a high-scoring sub network by the following steps:

- In the first step all positive connected nodes are aggregated into meta-nodes;
- By defining an edge score based on the node's scores that are on the endpoints of an edge, the node scores are transferred to the edges;
- On these edge scores a minimum spanning tree (MST) is calculated;
- All paths between positive meta-nodes are calculated based on the MST to obtain the negative nodes between the positives;
- Upon these negative nodes again a MST is calculated from which the path with the highest score, regarding node scores of negative nodes and the positive meta-nodes they connect, gives the resulting approximated module.

The result of the BioNet algorithm is a sub-network for each compound, or network module, that agglomerates highly significant nodes. The sizes of the modules are dependent on the algorithmic parameter, i.e. the false-discovery rate $t(\text{FDR})$. The computed modules are fairly stable given the algorithmic parameter. This is illustrated in Figure 5. Here, the BioNet approach was computed for rifampin with different thresholds ($\text{FDR} = 0.05, 0.02, 0.01$) resulting in gene lists of different sizes ($n = 44, 27, 15$). It can be seen in the VENN diagram that genes derived from more stringent parameters are mostly included in genes derived from less stringent parameters.

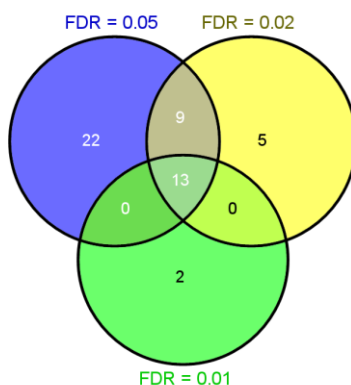


Figure 5: Consistency of BioNet output. The algorithm was run with different FDR thresholds yielding modules of different sizes that are highly overlapping.

1.7. Toxicity modules

Network modules were computed as described above for all five compounds. An overview representation of the key modules identified is displayed in Figure 6.

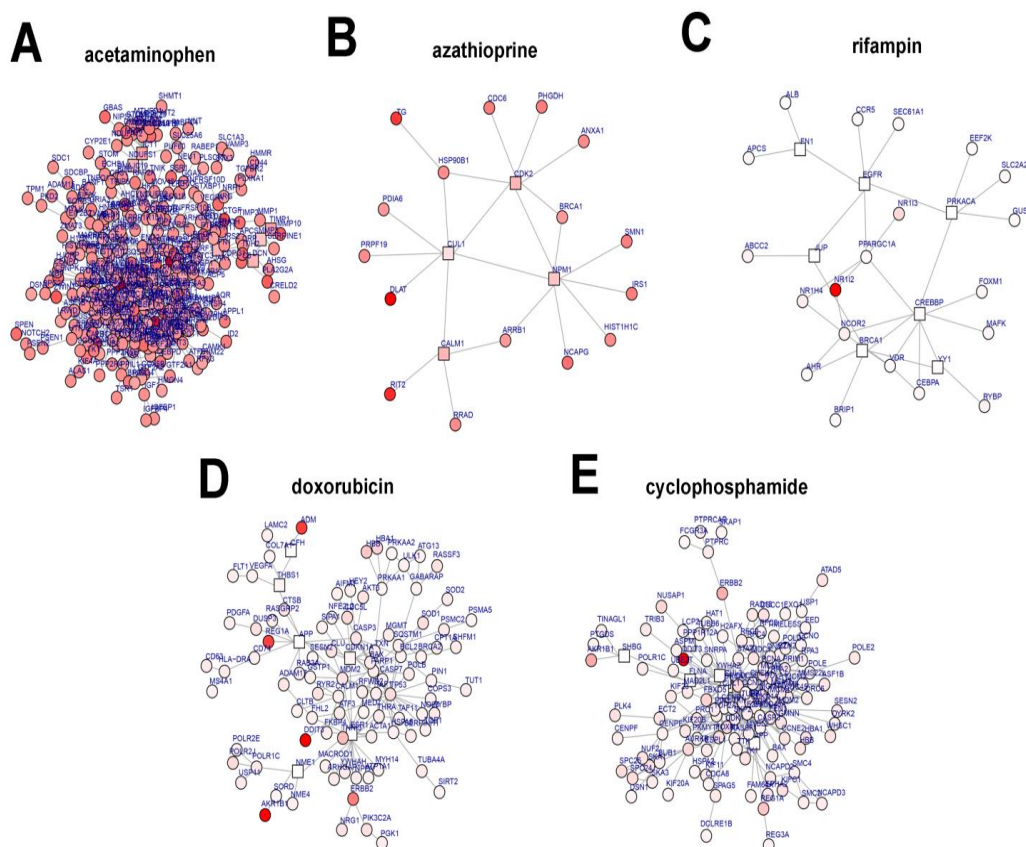


Figure 6: Combining PPI-information with the weighted top genes from literature mining and OMICS analysis key toxicity modules were computed using the BioNet algorithm ($t(FDR) = 0.02$). The five compounds acetaminophen, azathioprine, rifampin, doxorubicin and cyclophosphamide were analyzed. Node coloring represents the relative weight of each gene. Nodes in dark red are identified with very high significance and represent key mediators of toxicity.

Based on the different node weighing procedures derived from the p-values from literature mining and microarray analyses, the modules differ in gene content. Resulting gene lists contain 314 genes for acetaminophen, 20 for azathioprine, 27 for rifampin, 98 for doxorubicin and 125 for cyclophosphamide. The corresponding gene lists have been uploaded to the HeCaToS web site and are available for the project partners for further investigations.

We exemplify the results with the hepatotoxic compound rifampin. Rifampin is an antibiotic that inhibits DNA-dependent RNA polymerase activity in susceptible cells. Specifically, it interacts with bacterial RNA polymerase but does not inhibit the mammalian enzyme. It is bactericidal and has a very broad spectrum of activity against most gram-positive and gram-negative organisms and specifically *Mycobacterium tuberculosis*. Because of rapid emergence of resistant bacteria, use is restricted to treatment of mycobacterial infections and a few other indications. Rifampin is metabolized in the liver and eliminated in bile and, to a much lesser extent, in urine. The mechanism of rifampin hepatotoxicity is not well known, but it is extensively metabolized by the liver and induces multiple hepatic enzymes including CYP 3A4 and ABC C2 (MRP2). Thus, the cause of injury is likely to be due to idiosyncratic metabolic products that are either directly toxic or induce an immunologic reaction. The rise in direct and total bilirubin in rare patients receiving rifampin may relate to gene defects in MRP2 (ABC C2). Patients with pre-existing liver disease and cirrhosis are particularly likely to develop jaundice on rifampin therapy.

We uploaded the summarized p-values from literature mining and microarray analysis to the PPI network. The BioNet node weighting yields a transformation of the original p-value distribution into a Beta-distribution mixture (Figure 7A) which yield a reasonable fit (Figure 7B). Scoring with FDR threshold ($q = 0.02$) yields a module consisting of 27 genes (Table 2) that is highly responsive for rifampin treatment (Figure 7C).

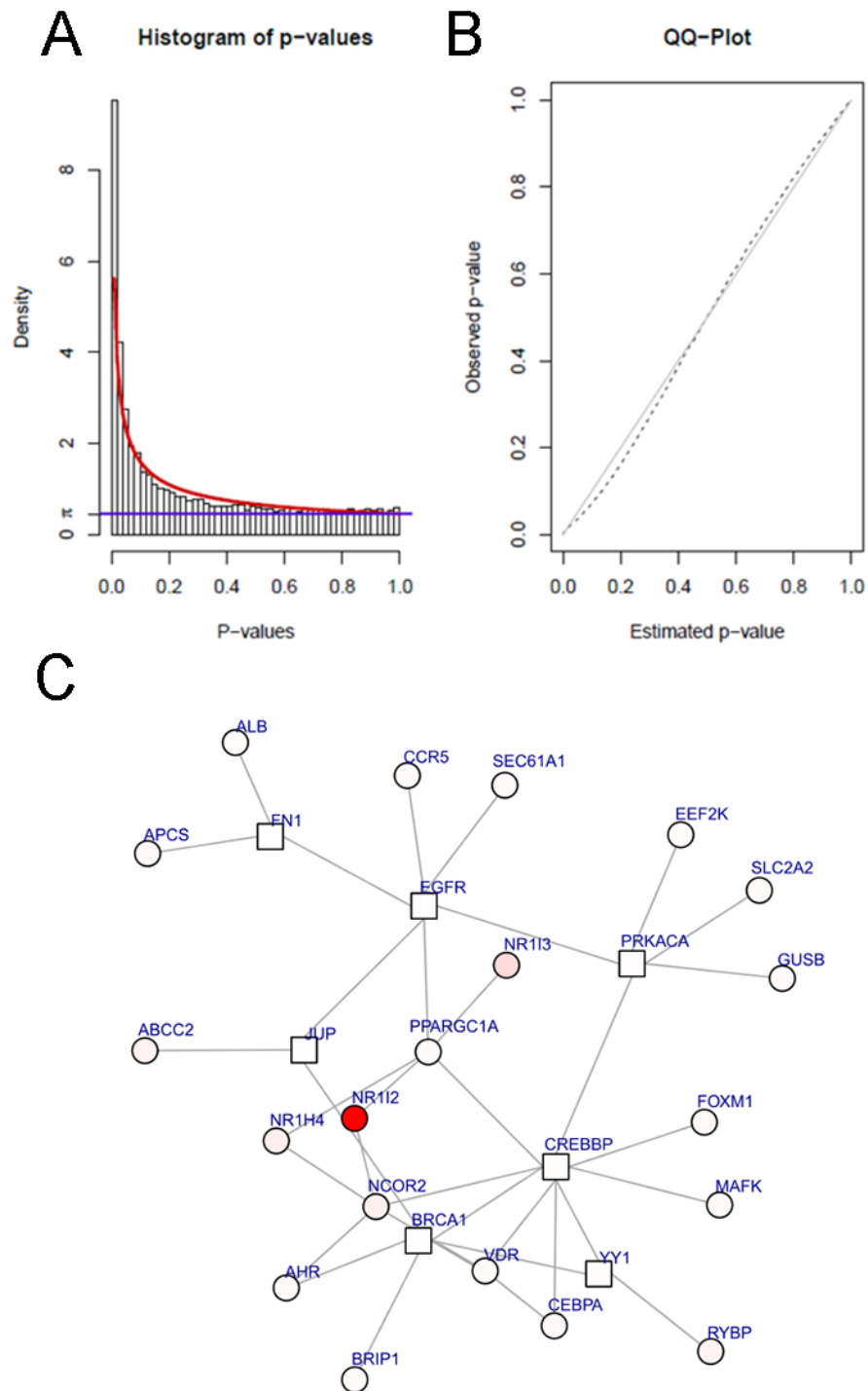


Figure 7: BioNet network module computation. A) Distribution of p-values from rifampin. Fitted curves correspond to the estimated Beta functions for signal and noise that are subsequently being used to weight the network nodes. B) QQ-plot of fitted (X-axis) vs. experimental (Y-axis) p-values. C) Computed module ($t(\text{FDR}) = 0.02$) with 27 response genes.

Table 5: Rifampin response genes.

BRCA1	RYBP	PRKACA
NCOR2	VDR	YY1
FOXM1	APCS	BRIP1
AHR	CEBPA	NR1H4
PPARGC1A	NR1I2	CCR5
EEF2K	MAFK	SLC2A2
SEC61A1	FN1	GUSB
EGFR	NR1I3	
JUP	CREBBP	
ALB	ABCC2	

The selected response genes enrich important pathways of rifampin metabolism and liver toxicity such as bile acid pathway, the role of nuclear receptors in lipid metabolism and toxicity (Figure 8). The different pathway concepts are highly overlapping as can be seen from Figure 9.

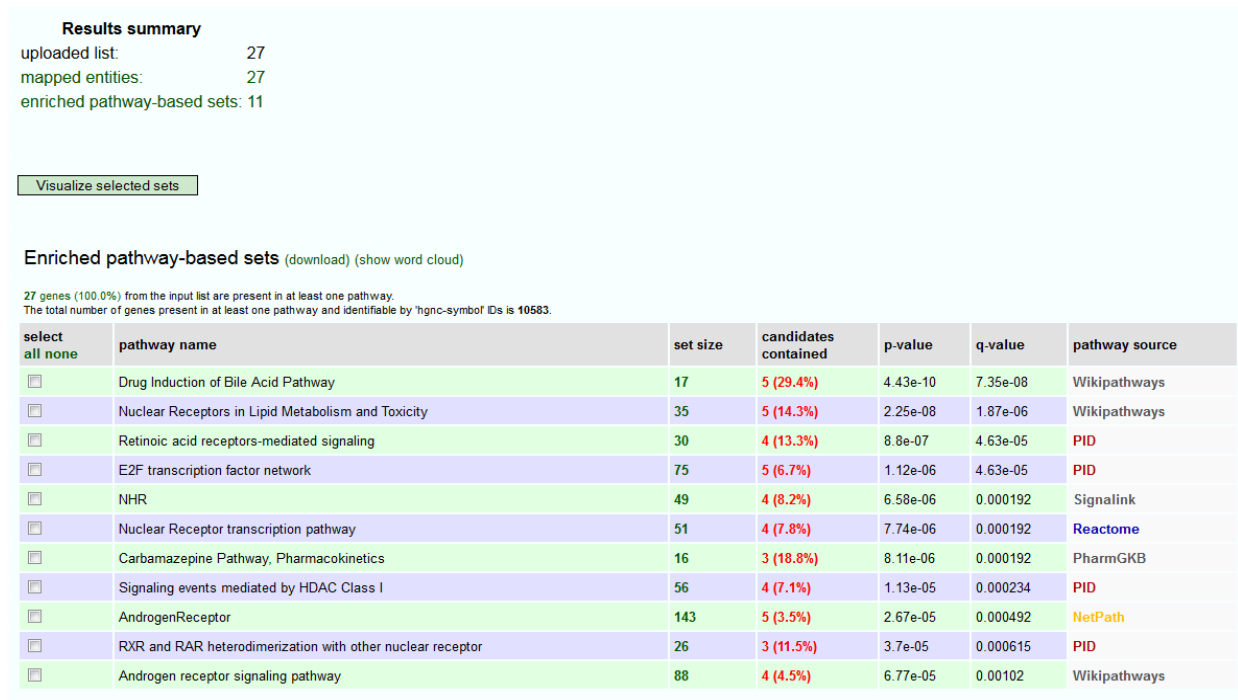


Figure 8: Over-representation analysis with module genes. Over-representation analysis was done within the ConsensusPathDB inserting the 27 proteins of the computed network module.

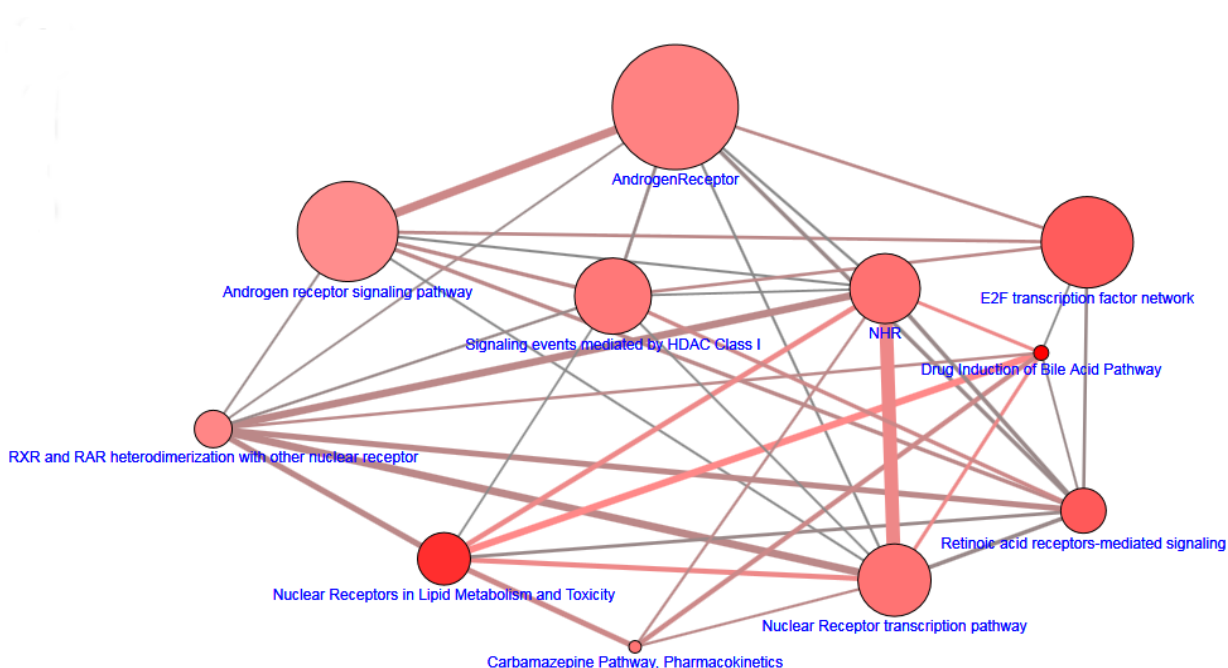


Figure 9: Visualization of enriched pathway concepts. Pathways are displayed as circles. Size of the pathway is reflected by different sizes of the circles. Overlap between pathway concepts are indicated by edges between circles. Thickness of edges indicate size of overlap.

Furthermore, the most reactive protein in the computed network module for rifampin is NR1I2. This gene product belongs to the nuclear receptor superfamily, members of which are transcription factors characterized by a ligand-binding domain and a DNA-binding domain. The encoded protein is a transcriptional regulator of the cytochrome P450 gene CYP3A4, binding to the response element of the CYP3A4 promoter as a heterodimer with the 9-cis retinoic acid receptor RXR. It is activated by a range of compounds that induce CYP3A4, including dexamethasone and rifampicin.

Thus, we conclude that the combined approach conducted here generates network modules that are predictive of compound toxicity. These modules will be further used for toxicity studies and construction of AOPs in **WPs 10** and **11** as well as initial interaction maps for molecular models developed further in **WP2**.

DIFFICULTIES

- The process of selecting appropriate targets for further investigation had to be settled among all parties and took some time;
- Prioritization sometimes leads to conflicts in the sense that either computational models or OMICS data may not be readily available;
- Results based on literature mining and results based on OMICS data may be rather different and sometimes difficult to match.

REFERENCES

- Hakenberg, J., Gerner, M., Haeussler, M., Solt, I., Plake, C., Schroeder, M., Gonzalez, G., Nenadic, G., and Bergman, C. M. (2011). The GNAT library for local and remote gene mention normalization. *Bioinformatics*, 27(19):2769-2771.
- Kamburov A, Grossmann A, Herwig R, Stelzl U (2012) Cluster-based assessment of protein-protein interaction confidence. *BMC Bioinformatics*, 13:262.
- Kamburov A, Stelzl U, Herwig R (2012) IntScore: a web tool for confidence scoring of biological interactions. *Nucleic Acids Research*, 40:W140-W146.
- Yamashita F, Sasa Y, Yoshida S, Hisaka A, Asai Y, Kitano H, Hashida M, Suzuki H (2013) Modeling of rifampin-induced CYP3A4 activation dynamics for the prediction of clinical drug-drug interactions from in vitro data. *PLoS ONE*, 8:e70330.

ANNEX: TOP-50 SCORING GENES PER COMPOUND DERIVED FROM A) LITERATURE MINING AND B) GENE EXPRESSION ANALYSIS

A) Literature mining

Acetaminophen (liver): top list of literature based genes

Symbol	Description	Syns	NTest	SizeTest	NRef	SizeRef	pv
RYBP	RING1 and YY1 binding protein	APAP(288)	288	2246	1052	12800000	0
GPT	glutamic-pyruvate transaminase (a	ALT(74);alanine aminotransferase(23	98	2246	15685	12800000	1,5984E-229
CYP2E1	cytochrome P450, family 2, subfam	CYP2E1(49);CYP2E(4);Cyp2e1(3);cyp2	59	2246	2692	12800000	3,5261E-204
SLC17A5	solute carrier family 17 (acidic suga	AST(22)	22	2246	3373	12800000	1,48955E-56
CYP1A2	cytochrome P450, family 1, subfam	CYP1A2(14);Cyp1a2(7)	21	2246	3190	12800000	2,08794E-54
UGT1A6	UDP glucuronosyltransferase 1 fam	UGT1A6(7);HlugP1(2);UGT 1A6(2)	11	2246	377	12800000	2,52847E-46
FMO3	flavin containing monooxygenase 3	Fmo3(9);flavin containing monooxyge	11	2246	460	12800000	2,74349E-44
SULT1A1	sulfotransferase family, cytosolic, 1	PST(10);phenol sulfotransferase(1);Su	12	2246	783	12800000	8,75637E-43
UGT1A1	UDP glucuronosyltransferase 1 fam	UDPGT(7);UGT1A1(7)	14	2246	1504	12800000	1,63943E-42
GCLM	glutamate-cysteine ligase, modifier	glutamate-cysteine ligase(7);GCL(2);g	11	2246	631	12800000	4,73818E-41
CYP3A4	cytochrome P450, family 3, subfam	CYP3A4(18);CYP3A(2)	20	2246	7081	12800000	2,92067E-37
F2	coagulation factor II (thrombin)	prothrombin(17)	17	2246	6149	12800000	4,5246E-32
UGT1A9	UDP glucuronosyltransferase 1 fam	UGT1A9(6);HlugP4(1);UGT 1A9(1)	8	2246	472	12800000	2,47316E-31
GC	group-specific component (vitamin	Gc-globulin(5);vitamin D-binding prot	9	2246	939	12800000	1,29129E-29
CYS1	cystin 1	CYS(6)	6	2246	199	12800000	1,79081E-28
RIPK3	receptor-interacting serine-threoni	RIP3(6)	6	2246	212	12800000	4,27145E-28
ABCC4	ATP-binding cassette, sub-family C	Mrp4(6);ABCC4(1);MRP4(1)	8	2246	953	12800000	6,12622E-26
ADO	2-aminoethanethiol (cysteamine) c	cysteamine(7);Cysteamine(1)	8	2246	1039	12800000	2,80653E-25
B3GAT3	beta-1,3-glucuronyltransferase 3 (g	glucuronosyltransferase(6)	6	2246	383	12800000	1,46597E-24
AADAC	arylacetamide deacetylase	AADAC(4)	4	2246	67	12800000	4,55219E-24
GABPA	GA binding protein transcription fa	GABP(6)	6	2246	461	12800000	1,88288E-23
COMA		coma(10);Coma(1)	11	2246	3768	12800000	6,1434E-23
HESX1	HESX homeobox 1	h-ANF(5)	5	2246	293	12800000	6,57041E-22
MIR561	microRNA 561	miR-561(2)	2	2246	4	12800000	1,16225E-20
MIR122	microRNA 122	miR-122(4);microRNA-122(1);mir-122	6	2246	819	12800000	5,06171E-20
NR1H4	nuclear receptor subfamily 1, group	FXR(8)	8	2246	2136	12800000	8,26094E-20
ALB	albumin	albumin(38);serum albumin(4);Album	44	2246	87358	12800000	2,56023E-19
SECISBP2	SECIS binding protein 2	SBP2(4)	4	2246	219	12800000	4,95709E-19
CCL4	chemokine (C-C motif) ligand 4	CCl4(7)	7	2246	1730	12800000	1,87878E-18
SLC13A1	solute carrier family 13 (sodium/su	Nas1(3);NaS1(2)	5	2246	590	12800000	2,53857E-18
CYP2B6	cytochrome P450, family 2, subfam	P450(8);P-450(2)	10	2246	4783	12800000	2,802E-18
ACOX1	acyl-CoA oxidase 1, palmitoyl	ACOX(4)	4	2246	314	12800000	1,72265E-17
NR1I2	nuclear receptor subfamily 1, group	PXR(4);SXR(4);Pxr(1)	9	2246	4034	12800000	2,06595E-17
SLCO1B3	solute carrier organic anion transp	OATP1B3(4);SLCO1B3(1)	5	2246	718	12800000	2,55448E-17
GSS	glutathione synthetase	glutathione synthetase(4)	4	2246	331	12800000	2,89481E-17
NFE2L2	nuclear factor, erythroid 2-like 2	Nrf2(8);Nrf-2(2)	10	2246	6783	12800000	3,20756E-15
EMP1	epithelial membrane protein 1	TMP(6)	6	2246	1889	12800000	4,2779E-15
MLKL	mixed lineage kinase domain-like	MLKL(2)	2	2246	47	12800000	9,75213E-15
TXNRD1	thioredoxin reductase 1	TrxR1(4);thioredoxin reductase(1)	5	2246	1232	12800000	1,40229E-14
SCLY	selenocysteine lyase	Scly(2);selenocysteine lyase(1)	3	2246	248	12800000	2,17138E-14
NR1I3	nuclear receptor subfamily 1, group	CAR(4);NR1I3(1)	5	2246	1317	12800000	3,03968E-14
PDIA3	protein disulfide isomerase family	microsomal(5)	5	2246	1444	12800000	8,81584E-14
DSTYK	dual serine/threonine and tyrosine	RIP5(1)	1	2246	4	12800000	9,45808E-14
UGP2	UDP-glucose pyrophosphorylase 2	UDPG(2)	2	2246	77	12800000	1,78251E-13
SULT1C4	sulfotransferase family, cytosolic, 1	SULT1C4(1)	1	2246	5	12800000	2,12757E-13
PRUNE2	prune homolog 2 (Drosophila)	BMCC(2)	2	2246	85	12800000	3,19571E-13
ABCC3	ATP-binding cassette, sub-family C	Mrp3(4)	4	2246	905	12800000	5,50475E-13
HRH2	histamine receptor H2	H2-receptor(2);histamine H2-recepto	3	2246	389	12800000	7,5168E-13
AFP	alpha-fetoprotein	AFP(5);alpha-fetoprotein(5)	10	2246	9055	12800000	8,94772E-13
CYP3A5	cytochrome P450, family 3, subfam	cytochrome P450(2);CYP3A5(1);cytod	5	2246	1767	12800000	8,99048E-13

Azathioprine (liver): top list of literature based genes

Symbol	Description	Syns	NTest	SizeTest	NRef	SizeRef	pv
TPMT	thiopurine S-methyltransferase	TPMT(35)	35	532	1840	12800000	4,5E-165
ITPA	inosine triphosphatase (nucleoside)	ITPA(6);inosine triphosphatase(1);i	9	532	461	12800000	3,34E-48
MTX1	metaxin 1	MTX(6)	6	532	1592	12800000	1,03E-24
EMP1	epithelial membrane protein 1	TMP(6)	6	532	1889	12800000	1,1E-23
DLAT	dihydrolipoamide S-acetyltransferase	PBC(3)	3	532	923	12800000	7,59E-15
CYP21A2	cytochrome P450, family 21, subfa	CAH(4)	4	532	2973	12800000	4,56E-14
TG	thyroglobulin	TGN(4)	4	532	3138	12800000	7,73E-14
RIT2	Ras-like without CAAX 2	RIBA(2)	2	532	300	12800000	1,03E-13
MTHFR	methylenetetrahydrofolate reductase	MTHFR(4);methylenetetrahydrofol	5	532	10082	12800000	2,64E-11
GPT	glutamic-pyruvate transaminase (a	ALT(3);alanine aminotransferase(2)	6	532	15685	12800000	2,97E-11
ACE	angiotensin I converting enzyme	ACE(8);angiotensin-converting enzy	9	532	39795	12800000	8E-11
RBM27	RNA binding motif protein 27	PSC(2)	2	532	920	12800000	8,16E-11
AMELX	amelogenin, X-linked	AIH(2)	2	532	995	12800000	1,3E-10
ATP2C1	ATPase, Ca++ transporting, type 2C	ATP2C1(2)	2	532	1105	12800000	2,42E-10
SMN1	survival of motor neuron 1, telome	SMA(3)	3	532	4050	12800000	8,4E-10
null		gout(3);Gout(1)	4	532	10320	12800000	6,98E-09
IL2	interleukin 2	IL-2(3);IL2(2);interleukin-2(2);Interl	8	532	44255	12800000	1,52E-08
FUT3	fucosyltransferase 3 (galactoside 3	LEs(1)	1	532	535	12800000	5,93E-08
ZIC3	Zic family member 3	HTX(1)	1	532	736	12800000	2,1E-07
PKD1P5	polycystic kidney disease 1 (autoso	HGV(1)	1	532	778	12800000	2,61E-07
BLOC1S1	biogenesis of lysosomal organelles	bloc(1)	1	532	833	12800000	3,43E-07
AA1		alopecia areata(1)	1	532	1026	12800000	7,8E-07
GGT1	gamma-glutamyltransferase 1	GGT(2)	2	532	4839	12800000	1,35E-06
SHFM1	split hand/foot malformation (ectro	SEM(3)	3	532	11200	12800000	1,8E-06
EFTUD1	elongation factor Tu GTP binding d	RIA(1)	1	532	1444	12800000	2,99E-06
SOX10	SRY (sex determining region Y)-box	Dom(1)	1	532	1647	12800000	5E-06
PART1	prostate androgen-regulated trans	part I(1)	1	532	1872	12800000	8,24E-06
TCF15	transcription factor 15 (basic helix-	EC 2(1)	1	532	1876	12800000	8,31E-06
PHGDH	phosphoglycerate dehydrogenase	Sera(1)	1	532	2015	12800000	1,1E-05
CYP3A4	cytochrome P450, family 3, subfam	CYP3A4(2)	2	532	7081	12800000	1,16E-05
ABCD1	ATP-binding cassette, sub-family D	ALD(1)	1	532	2459	12800000	2,38E-05
IRS1	insulin receptor substrate 1	IRS-1(1)	1	532	2789	12800000	3,86E-05
IL23A	interleukin 23, alpha subunit p19	IL-23(1)	1	532	2884	12800000	4,39E-05
MAGED2	melanoma antigen family D, 2	BCG(1)	1	532	3072	12800000	5,59E-05
MMP1	matrix metalloproteinase 1 (intersti	MMP(3)	3	532	18950	12800000	7,32E-05
SLC17A5	solute carrier family 17 (acidic suga	AST(1)	1	532	3373	12800000	8E-05
RPS6KB1	ribosomal protein S6 kinase, 70kDa	p70S6K(1)	1	532	3756	12800000	0,00012
COMA		Coma(1)	1	532	3768	12800000	0,000122
LDHA	lactate dehydrogenase A	LDH(1)	1	532	4522	12800000	0,000243
TSC2	tuberous sclerosis 2	TSC2(1)	1	532	4630	12800000	0,000265
IFNB1	interferon, beta 1, fibroblast	interferon-beta(1)	1	532	5392	12800000	0,000468
EPHB2	EPH receptor B2	ERK(1)	1	532	5649	12800000	0,000556
RAC1	ras-related C3 botulinum toxin sub	Rac1(1)	1	532	6591	12800000	0,000979
IL17A	interleukin 17A	IL-17(1)	1	532	7601	12800000	0,00164
CD40LG	CD40 ligand	IgM(1)	1	532	13162	12800000	0,010966
ICAM1	intercellular adhesion molecule 1	intercellular adhesion molecule-1(1	1	532	25004	12800000	0,077664
REN	renin	renin(1)	1	532	25602	12800000	0,082842
EPO	erythropoietin	Erythropoietin(1)	1	532	27486	12800000	0,10018
ABCB1	ATP-binding cassette, sub-family B	P-glycoprotein(1)	1	532	42427	12800000	0,27726
TNF	tumor necrosis factor	TNF(4);TNFalpha(1)	5	532	180907	12800000	0,581515

Rifampin (liver): top list of literature based genes

Symbol	Description	Syns	NTest	SizeTest	NRef	SizeRef	pv
CYP3A4	cytochrome P450, family 3, subfam	CYP3A4(94);CYP3A(24);cytochrome	122	783	7081	12800000	0
NR1I2	nuclear receptor subfamily 1, group	PXR(98);SXR(9);NR1I2(3);pxr(3);Pxr	115	783	4034	12800000	0
AADAC	arylacetamide deacetylase	AADAC(14);arylacetamide deacetyl	15	783	67	12800000	1,9E-102
SLCO1B1	solute carrier organic anion transp	OATP1B1(12);OATP-C(3);SLCO1B1(21	783	1468	12800000	4,43E-89
NR1I3	nuclear receptor subfamily 1, group	CAR(17);NR1I3(1)	18	783	1317	12800000	1,31E-76
SLC10A1	solute carrier family 10 (sodium/bi	NTCP(11);SLC10A1(2)	13	783	532	12800000	3,07E-64
ABCB11	ATP-binding cassette, sub-family B	BSEP(11);ABCB11(1);ABCB 11(1);bi	15	783	1339	12800000	2,9E-62
SLCO1B3	solute carrier organic anion transp	OATP8(6);OATP1B3(4);organic anio	12	783	718	12800000	1,24E-55
GPT	glutamic-pyruvate transaminase (a	ALT(20);alanine aminotransferase(24	783	15685	12800000	4,21E-53
G6PC	glucose-6-phosphatase, catalytic su	G6Pase(13);glucose 6-phosphatase	15	783	2808	12800000	4,38E-52
CYP2B6	cytochrome P450, family 2, subfam	CYP2B6(8);P450(7);CYP2B(1)	16	783	4783	12800000	2,44E-48
MIR34A	microRNA 34a	miR-34a(11)	11	783	1162	12800000	1E-45
CYP2E1	cytochrome P450, family 2, subfam	CYP2E1(11);cytochrome P-4502E1(12	783	2692	12800000	7,09E-41
CYP1A2	cytochrome P450, family 1, subfam	CYP1A2(12)	12	783	3190	12800000	5,49E-39
SLCO1A2	solute carrier organic anion transp	OATP(4);Oatp(3);organic anion tran	8	783	516	12800000	7,27E-39
SGK3	serum/glucocorticoid regulated kin	SGK2(7)	7	783	274	12800000	2,64E-38
CYP3A5	cytochrome P450, family 3, subfam	cytochrome P450(5);CYP3A5(4);cyt	10	783	1767	12800000	2,64E-37
NR1H4	nuclear receptor subfamily 1, group	FXR(10)	10	783	2136	12800000	1,63E-35
NCOR2	nuclear receptor corepressor 2	SMRT(8)	8	783	984	12800000	7,19E-34
ABCC3	ATP-binding cassette, sub-family C	MRP3(6);Abcc3(1)	7	783	905	12800000	4,3E-30
GSTT1	glutathione S-transferase theta 1	GSTT1(9)	9	783	3784	12800000	8,91E-27
RYBP	RING1 and YY1 binding protein	APAP(6)	6	783	1052	12800000	7,25E-25
SLC17A5	solute carrier family 17 (acidic suga	AST(8)	8	783	3373	12800000	2,25E-24
CYP2A6	cytochrome P450, family 2, subfam	CYP2A6(6)	6	783	1149	12800000	2,46E-24
ABCC2	ATP-binding cassette, sub-family C	mrp2(2);MRP2(2);Abcc2(1);multidr	7	783	2738	12800000	1,67E-22
CES4A	carboxylesterase 4A	Carboxylesterase 6(1);Ces6(1)	2	783	8	12800000	7,48E-22
AMCN		AMC(4)	4	783	652	12800000	6,65E-19
MAFK	v-maf avian musculoaponeurotic fi	MAFK(2)	2	783	69	12800000	1,69E-16
GSTM1	glutathione S-transferase mu 1	GSTM1(5)	5	783	2545	12800000	2,92E-16
MRPS2	mitochondrial ribosomal protein S2	MRPs 2(1)	1	783	4	12800000	1,4E-15
NAT2	N-acetyltransferase 2 (arylamine N	NAT2(4)	4	783	1758	12800000	1,18E-14
CEBPA	CCAAT/enhancer binding protein (C	C/EBPalpha(4);CCAAT/enhancer bin	5	783	3515	12800000	1,27E-14
CYP2C8	cytochrome P450, family 2, subfam	CYP2C8(3)	3	783	683	12800000	1,51E-14
CYP3A7	cytochrome P450, family 3, subfam	CYP3A7(2)	2	783	301	12800000	1,07E-12
TTPA	tocopherol (alpha) transfer protein	tocopherol(4);Tocopherol(1)	5	783	5925	12800000	5,18E-12
AHR	aryl hydrocarbon receptor	AhR(6)	6	783	10007	12800000	1,35E-11
SLC5A8	solute carrier family 5 (sodium/mo	AIT(2)	2	783	480	12800000	1,71E-11
BRIP1	BRCA1 interacting protein C-termi	BACH1(2)	2	783	500	12800000	2,18E-11
PPARGC1A	peroxisome proliferator-activated r	PGC-1alpha(3)	3	783	2047	12800000	8,52E-11
WSB1	WD repeat and SOCS box containin	WSB(1)	1	783	88	12800000	2,13E-10
GGT1	gamma-glutamyltransferase 1	GGT(4)	4	783	4839	12800000	2,15E-10
CYP2C19	cytochrome P450, family 2, subfam	CYP2C19(3)	3	783	2362	12800000	2,59E-10
APCS	amyloid P component, serum	SAP(2)	2	783	837	12800000	4,63E-10
ATP6AP1	ATPase, H+ transporting, lysosoma	16alpha(2)	2	783	916	12800000	7,9E-10
FOXM1	forkhead box M1	HNF3(2)	2	783	1195	12800000	3,79E-09
CES2	carboxylesterase 2	CES2(1)	1	783	184	12800000	3,99E-09
SEC61A1	Sec61 alpha 1 subunit (S. cerevisiae	Sec61(1)	1	783	209	12800000	6,61E-09
PPP2R4	protein phosphatase 2A activator,	PP2A(3)	3	783	3714	12800000	8,49E-09
CCR5	chemokine (C-C motif) receptor 5	CCR5(5)	5	783	11663	12800000	9,68E-09
VDR	vitamin D (1,25- dihydroxyvitamin	VDR(5)	5	783	12076	12800000	1,41E-08

Doxorubicin (heart): top list of literature based genes

Symbol	Description	Syns	NTest	SizeTest	NRef	SizeRef	pv
AKR1B1	aldo-keto reductase family 1, mem	ADR(31);Adr(7)	38	2771	2660	12800000	1,6E-112
DDIT3	DNA-damage-inducible transcript 3	CHOP(53)	53	2771	9417	12800000	4,4E-112
REG1A	regenerating islet-derived 1 alpha	ICRF(25)	25	2771	1167	12800000	4,39E-85
ADM	adrenomedullin	ADM(29);adrenomedullin(2)	31	2771	2897	12800000	4,53E-85
ERBB2	v-erb-b2 avian erythroblastic leuke	HER2(33);ErbB2(5);HER2/neu(5);H	50	2771	28130	12800000	1,53E-58
ABCB1	ATP-binding cassette, sub-family B	P-glycoprotein(29);P-gp(9);Pgp(5);F	52	2771	42427	12800000	5,35E-46
CBR1	carbonyl reductase 1	CBR1(4);HCBR(4);Carbonyl reducta	9	2771	285	12800000	5,72E-38
CSF3	colony stimulating factor 3 (granulo	granulocyte colony-stimulating fact	17	2771	3959	12800000	1,81E-35
MYOCD	myocardin	myocardin(8);Myocardin(1)	9	2771	973	12800000	1,62E-27
USP9X	ubiquitin specific peptidase 9, X-lin	FAM(8)	8	2771	750	12800000	3,73E-26
CYP2J2	cytochrome P450, family 2, subfam	CYP2J2(6)	6	2771	280	12800000	3,65E-25
TNNI3	troponin I type 3 (cardiac)	cTnl(5);cardiac troponin T(4);cardia	13	2771	4456	12800000	7,79E-24
HN1	hematological and neurological exp	arm 2(5)	5	2771	253	12800000	1,42E-21
RETN	resistin	resistin(9);Retn(2)	11	2771	3552	12800000	1,91E-21
AKR1A1	aldo-keto reductase family 1, mem	AKR1A1(4)	4	2771	102	12800000	2,22E-21
HOPX	HOP homeobox	HOP(5);homeodomain only protein	6	2771	541	12800000	3,11E-21
UGT2B7	UDP glucuronosyltransferase 2 fam	UGT2B7(6)	6	2771	595	12800000	1,15E-20
ABCC1	ATP-binding cassette, sub-family C	multidrug resistance(12);MDR(2);M	15	2771	8621	12800000	3,09E-20
EGLN3	egl-9 family hypoxia-inducible fact	PHD3(5)	5	2771	375	12800000	1,47E-19
EDNRA	endothelin receptor type A	ETA(5);EDNRA(2)	7	2771	1196	12800000	1,62E-19
CKMT2	creatine kinase, mitochondrial 2 (sa	sMTCK(3)	3	2771	45	12800000	1,79E-19
ACO1	aconitase 1, soluble	IRP1(4);IRP-1(2)	6	2771	880	12800000	2,41E-18
CDC5L	cell division cycle 5-like	CEF(5)	5	2771	504	12800000	4,77E-18
TF	transferrin	transferrin(15);TRF(2)	17	2771	15030	12800000	1,39E-16
TP53	tumor protein p53	p53(79);p53 tumor suppressor(1)	80	2771	188350	12800000	1,88E-16
EPHX2	epoxide hydrolase 2, cytoplasmic	sEH(5)	5	2771	730	12800000	3,67E-16
RGS6	regulator of G-protein signaling 6	RGS6(3)	3	2771	129	12800000	6,7E-16
POLR2J	polymerase (RNA) II (DNA directed	RPB11(2)	2	2771	28	12800000	1,68E-15
CCNO	cyclin O	CCNU(3)	3	2771	159	12800000	3,47E-15
KRTAP5-9	keratin associated protein 5-9	KRN(2)	2	2771	33	12800000	4,35E-15
DUSP3	dual specificity phosphatase 3	VHR(3)	3	2771	204	12800000	2,47E-14
CTSB	cathepsin B	cathepsin B(4)	4	2771	576	12800000	5,26E-14
GLRX2	glutaredoxin 2	Glrx2(3)	3	2771	227	12800000	5,72E-14
MPPED1	metallophosphoesterase domain c	FAMia(1)	1	2771	3	12800000	7,89E-14
AKT3	v-akt murine thymoma viral oncoge	Akt3(4)	4	2771	709	12800000	3,98E-13
FECH	ferrochelataase	FCE(5)	5	2771	1445	12800000	1E-12
MB	myoglobin	myoglobin(7)	7	2771	3367	12800000	1,07E-12
TERF2	telomeric repeat binding factor 2	TRF2(3);telomeric repeat binding fa	4	2771	804	12800000	1,35E-12
CYP27A1	cytochrome P450, family 27, subfa	CTX(4)	4	2771	824	12800000	1,71E-12
EIF2C2		Q10(6)	6	2771	2416	12800000	1,82E-12
ERAL1	Era-like 12S mitochondrial rRNA ch	HERA(2)	2	2771	97	12800000	2,45E-12
PBCA		PBCA(2)	2	2771	112	12800000	5,72E-12
ZMAT3	zinc finger, matrin-type 3	wig(2)	2	2771	126	12800000	1,15E-11
MZT1	mitotic spindle organizing protein 1	MZT(2)	2	2771	130	12800000	1,38E-11
RYBP	RING1 and YY1 binding protein	APAP(4)	4	2771	1052	12800000	1,8E-11
CASP3	caspase 3, apoptosis-related cystei	caspase-3(22);Caspase-3(2);caspas	25	2771	42618	12800000	1,87E-11
AKR1C3	aldo-keto reductase family 1, mem	AKR1C3(4)	4	2771	1083	12800000	2,38E-11
NRG1	neuregulin 1	neuregulin-1(3);NRG1(2);neureguli	6	2771	3028	12800000	3,4E-11
CCNDBP1	cyclin D-type binding-protein 1	MAID(2)	2	2771	158	12800000	4,38E-11
PIK3C2A	phosphatidylinositol-4-phosphate 3	CPK(4)	4	2771	1172	12800000	5,08E-11

Cyclophosphamide (heart): top list of literature based genes

Symbol	Description	Syns	NTest	SizeTest	NRef	SizeRef	pv
DDIT3	DNA-damage-inducible transcript 3	CHOP(50)	50	948	9417	12800000	1E-150
AKR1B1	aldo-keto reductase family 1, mem	ADR(14);Adr(1)	15	948	2660	12800000	3,56E-50
ERBB2	v-erb-b2 avian erythroblastic leuke	HER2(21);HER-2(3);HER2/neu(3);Er	29	948	28130	12800000	1,56E-48
REG1A	regenerating islet-derived 1 alpha	ICRF(9)	9	948	1167	12800000	3,45E-35
CSF3	colony stimulating factor 3 (granulo	granulocyte colony-stimulating fact	11	948	3959	12800000	3,74E-31
CPA1	carboxypeptidase A1 (pancreatic)	CPA(9)	9	948	2346	12800000	3,3E-29
IL2	interleukin 2	IL-2(14);interleukin-2(3);IL2(2);Inte	22	948	44255	12800000	1E-24
CDC5L	cell division cycle 5-like	CEF(5)	5	948	504	12800000	1,36E-23
HN1	hematological and neurological exp	arm 2(4)	4	948	253	12800000	3,8E-22
NPHS1	nephrosis 1, congenital, Finnish typ	CNF(6)	6	948	1526	12800000	1,76E-21
IFNB1	interferon, beta 1, fibroblast	IFNbeta(8)	8	948	5392	12800000	2,33E-19
CYP27A1	cytochrome P450, family 27, subfa	CTX(4)	4	948	824	12800000	4,51E-17
OAP		OAP(2)	2	948	78	12800000	1,1E-15
COL18A1	collagen, type XVIII, alpha 1	endostatin(5)	5	948	2489	12800000	2,13E-15
ERAL1	Era-like 12S mitochondrial rRNA ch	HERA(2)	2	948	97	12800000	3,99E-15
MZT1	mitotic spindle organizing protein 1	MZT(2)	2	948	130	12800000	2,27E-14
CCNDBP1	cyclin D-type binding-protein 1	MAID(2)	2	948	158	12800000	7,23E-14
TNNT2	troponin T type 2 (cardiac)	cTnT(1);TNNT2(1);troponin T type 2	3	948	729	12800000	1,15E-13
ANC		ANC(3)	3	948	843	12800000	3,63E-13
EFS	embryonal Fyn-associated substrat	EFS(3)	3	948	852	12800000	3,95E-13
HLA-DRB4	major histocompatibility complex, I	Leukocyte Antigen(3)	3	948	1093	12800000	2,81E-12
POLR1C	polymerase (RNA) I polypeptide C,	AC 40(1)	1	948	25	12800000	3,16E-12
STEAP2	STEAP family member 2, metallo	STAMP I(1)	1	948	32	12800000	8,34E-12
ADM	adrenomedullin	ADM(4)	4	948	2897	12800000	9,92E-12
PROM1	prominin 1	hematopoietic stem cell(6);Hemat	7	948	11571	12800000	1,09E-11
OR13F1	olfactory receptor, family 13, subfa	or 96(1)	1	948	36	12800000	1,33E-11
PIN1	peptidylprolyl cis/trans isomerase,	DOD(3)	3	948	1524	12800000	3,8E-11
VDI		VDI(1)	1	948	70	12800000	1,84E-10
COMP	cartilage oligomeric matrix protein	COMP(3)	3	948	1983	12800000	2,95E-10
ZMYND10	zinc finger, MYND-type containing	FLU(2)	2	948	705	12800000	5,22E-10
PRTN3	proteinase 3	PR3(3);cANCA(1)	4	948	4903	12800000	1,49E-09
FCGR2A	Fc fragment of IgG, low affinity IIa,	FCGR2A(2)	2	948	977	12800000	3,58E-09
OR1B1	olfactory receptor, family 1, subfan	or 1b(1)	1	948	152	12800000	4E-09
MUTED		muted(1)	1	948	156	12800000	4,43E-09
CCNO	cyclin O	CCNU(1)	1	948	159	12800000	4,78E-09
CTC1	CTS telomere maintenance comple	CTC(2)	2	948	1053	12800000	5,56E-09
FABP3	fatty acid binding protein 3, muscle	heart fatty acid-binding protein(1);	2	948	1089	12800000	6,77E-09
FCGR3A	Fc fragment of IgG, low affinity IIIa,	FCGR3A(1);FCGR3A(1)	2	948	1123	12800000	8,11E-09
IS5		is 5(2)	2	948	1381	12800000	2,72E-08
RTE1		rte(1)	1	948	265	12800000	3,63E-08
EXO	endo/exonuclease (5'-3'), endonuc	Engl(1)	1	948	266	12800000	3,69E-08
FRA1L		frail(2)	2	948	1534	12800000	5,03E-08
CYTIP	cytohesin 1 interacting protein	B31(1)	1	948	300	12800000	5,94E-08
ITGA1	integrin, alpha 1	VLA(1)	1	948	300	12800000	5,94E-08
TNNI1	troponin I type 1 (skeletal, slow)	troponin I(2)	2	948	1636	12800000	7,32E-08
VEGFA	vascular endothelial growth factor	VEGF(18);vascular endothelial grow	22	948	135075	12800000	7,43E-08
SPPL2B	signal peptide peptidase like 2B	PSL(1)	1	948	365	12800000	1,29E-07
TNNI3	troponin I type 3 (cardiac)	cTnI(3)	3	948	4456	12800000	1,43E-07
EED	embryonic ectoderm development	heed(1)	1	948	412	12800000	2,09E-07
AQP4	aquaporin 4	AQP4(1);aquaporin-4(1)	2	948	2010	12800000	2,41E-07

B) Microarray analysis

Acetaminophen (liver): top list of differentially expressed genes

geneid	logfc	pval
HIST2H2BE	3,576	6,13E-07
GGTLC1	-1,997	1,09E-06
TXNRD1	1,594	1,95E-06
PIGH	1,376	1,08E-05
GPAM	-2,818	1,34E-05
HSD3B7	-1,308	1,93E-05
PEG10	3,132	2,09E-05
LEAP2	-2,845	2,14E-05
COL5A2	-1,29	2,34E-05
CTGF	1,602	2,65E-05
PRC1	-4,857	2,95E-05
GNA13	1,222	3,86E-05
MMP10	2,684	4,85E-05
NAT8	-5,012	5,01E-05
SLC6A6	1,863	6,03E-05
CEP57	-1,775	6,51E-05
HSBP1L1	-1,653	8,15E-05
ZC3HAV1	-1,823	9,5E-05
SPAG4	-1,338	0,000109
ZWINT	-1,933	0,00011
CYP4F11	1,249	0,000117
C4orf19	-1,856	0,000118
KMO	-4,325	0,000119
HRG	-1,631	0,000125
ZNF280C	1,429	0,000129
SHCBP1	-4,947	0,000135
HIST1H2BG	4,777	0,000145
TOP2A	-5,653	0,000151
F5	-1,166	0,000168
GJB1	-2,752	0,000169
SLC17A9	-1,974	0,000176
SH3BP5	1,882	0,000176
HGD	-1,599	0,000189
CYP2U1	1,14	0,00019
SLC25A18	-2,043	0,000192
IDI1	1,091	0,000196
ZSWIM5	1,182	0,000209
CENPK	-3,017	0,000215
FBXO16	-2,211	0,000228
MCM7	-2,645	0,00023
SLC41A2	1,01	0,000231
CLRN3	-6,072	0,000232
ARL15	1,96	0,000234
PDZK1IP1	-4,225	0,000238
ZNF512B	-1,002	0,000242

MND1	-3,217	0,000244
CYP8B1	-4,68	0,000251
NNMT	-2,116	0,000284
TIMP3	1,624	0,000289
SLC51A	-3,845	0,000304

Azathioprine (liver): top list of differentially expressed genes

geneid	logfc	pval
FEM1C	1,53	1,13E-05
TESC	-2,498	1,53E-05
PHF19	-1,265	2,15E-05
NCAPG	-2,67	2,31E-05
CRELD2	-2,043	2,39E-05
KIF4A	-3,989	2,62E-05
CDC6	-3,75	4,93E-05
PBK	-3,532	5,99E-05
RNASEH2A	-1,78	7,33E-05
POR	1,171	8,35E-05
BORA	-1,442	9,26E-05
HSPA1A	2,317	9,4E-05
ANXA1	1,046	0,000132
DMKN	-1,44	0,000147
NID1	-1,213	0,000152
HIST1H1C	1,007	0,000162
IRF2BPL	-1,039	0,000207
HSP90B1	-1,17	0,000236
BCHE	-1,026	0,000276
KIF18B	-3,315	0,000301
SLC10A1	-2,395	0,000313
RBP5	-1,403	0,000326
MBL2	-1,496	0,00034
GPX2	1,142	0,000349
BUB1	-1,817	0,000349
SDF2L1	-2,442	0,000361
C10orf10	1,121	0,000373
TYMS	-1,543	0,000377
ANKRD33	1,119	0,000379
TRAF1	-1,158	0,000428
HRG	-1,234	0,000433
SPC25	-3,512	0,000436
KLF11	1,149	0,000438
FMO1	-2,318	0,000441
NEK2	-2,811	0,000489
PIR	1,573	0,001
CYP4F11	1,187	0,001
SLC7A9	-1,191	0,001
ALDH8A1	-1,463	0,001
TTK	-3,846	0,001

MAFF	1,447	0,001
GAS2L3	-2,663	0,001
EFEMP1	-1,667	0,001
PRC1	-3,936	0,001
HLX	1,667	0,001
SLC25A30	1,245	0,001
CIDEA	1,055	0,001
BCL2A1	-4,267	0,001
C5	-1,356	0,001
MYBL2	-1,357	0,001

Rifampin (liver): top list of differentially expressed genes

geneid	logfc	pval
SERPINB9	1,271	1,43E-05
CYP2C9	1,419	1,59E-05
CYP3A4	7,875	0,000118
CYP3A7	6,215	0,000228
CXCL2	-1,262	0,000283
SEC14L4	1,186	0,001
VLDLR	1,346	0,001
SLC27A3	-1,764	0,001
CYP2C8	1,785	0,001
ALAS1	1,883	0,001
SLC7A5	1,007	0,001
THRSP	4,236	0,001
TSKU	3,151	0,001
CYP2C19	1,078	0,002
PRODH2	1,097	0,002
RRAD	-1,129	0,002
UPP1	1,135	0,002
DIO1	1,595	0,002
RHOF	1,189	0,002
INSIG1	1,24	0,002
CYP3A5	1,97	0,002
C19orf80	2,083	0,003
CD14	1,091	0,003
CYP3A43	1,72	0,003
IGF1	-1,173	0,003
MBL2	1,285	0,004
ADRB1	1,281	0,004
POR	1,778	0,004
STC2	2,144	0,004
TLR1	1,575	0,004
CTH	1,079	0,005
KMO	-1,106	0,005
ETNPPL	1,585	0,005
CYP2B6	2,713	0,006
SGK2	1,286	0,006

SLC10A1	-1,646	0,007
CYP2B7P	2,144	0,007
SLC7A11	1,105	0,009
FGF21	1,462	0,009
F13B	1,304	0,011
SCD	1,649	0,012
IFIT1	-1,405	0,013
GPAM	1,079	0,013
SERPINA7	-1,273	0,013
STK39	2,342	0,015
SLC2A1	1,27	0,017
PYCARD	-1,28	0,018
PCSK9	1,032	0,019
CMPK2	-1,228	0,02
PTPRJ	1,038	0,021

Doxorubicin (heart): top list of differentially expressed genes

geneid	logfc	pval
ALAS2	-6,05	1,82E-08
TRAK1	-2,23	5,15E-08
CLU	-2,63	4,12E-07
TFRC	-4,12	4,31E-07
RYS2	-1,98	5,43E-07
PTGR1	2,09	6,07E-07
BEX2	4	8,03E-07
C17orf58	1,5	1,69E-06
AGPAT3	-1,14	3,65E-06
MDK	1,15	4,08E-06
UCP2	-1,9	5,4E-06
KLHDC8B	1,51	6,32E-06
EPHX1	3,39	7,24E-06
MYH7	3,65	9,05E-06
GSTZ1	2,59	1,21E-05
BLOC1S6	1,31	1,26E-05
GIN1	1,76	0,000017
PPAPDC3	-1,14	1,85E-05
CDKN1C	1,23	2,19E-05
CPT1A	1,36	0,000024
FAM46C	-2,82	2,83E-05
SAT2	1,03	0,000033
SLC2A4	-1,83	3,59E-05
SESN1	1,31	0,000041
LIMS2	-1,9	4,29E-05
HLA-DRA	-2,61	4,29E-05
WBP5	1,67	4,45E-05
YWHAH	-1,12	4,48E-05
AKIP1	1,76	4,76E-05
SNRNP25	1,06	6,87E-05

SERPINH1	-1,2	7,54E-05
CXCL11	2,22	8,69E-05
APRT	-1,05	9,79E-05
IGKV3-11	-1,18	0,000105
IGKC	-1,18	0,000105
IGKV1-8	-1,18	0,000105
ALAS1	-1,09	0,000117
VAMP8	1,32	0,000119
SBSN	1,6	0,000123
GLUL	1,15	0,000161
EMP2	1,26	0,000165
CSF1R	-1,27	0,000172
LRRC48	1,01	0,000178
MTFP1	-1,77	0,000178
TMEM176B	-1,09	0,000204
PDP2	-1,35	0,000214
SLC27A1	1,22	0,000221
PLA2G5	-1,71	0,000248
AQP7	-1,23	0,000253
SQSTM1	1,13	0,000254

Cyclophosphamide (heart): top list of differentially expressed genes

geneid	logfc	pval
PBK	3,16	2,1E-14
ALAS2	-5,3	4,59E-14
FBXO5	3,3	1,26E-13
MELK	2,14	5,37E-13
NUSAP1	2,31	5,95E-13
CDK1	3,54	2,98E-12
AURKB	2,31	8,01E-12
KIF22	2,57	4,35E-11
SPC25	2,48	9,71E-11
KIAA0101	5,13	1,24E-10
FOXN1	1,08	1,26E-10
SPC24	1,79	2,09E-10
SHCBP1	2,19	2,13E-10
CDKN2C	1,75	2,29E-10
MGMT	1,59	2,35E-10
SKA1	1,82	7,07E-10
CDT1	1,8	8,08E-10
KIFC1	1,4	8,23E-10
CDKN3	2,79	2,04E-09
PLK4	2,48	2,95E-09
MDC1	1,17	4,55E-09
DSCC1	1,54	4,82E-09
ATAD5	1,93	5,46E-09
CCNE2	3,05	6,2E-09
ECT2	2,15	7,19E-09

IQGAP3	1,27	8,24E-09
WHSC1	1,34	8,75E-09
NUF2	1,83	1,08E-08
GMNN	1,18	1,09E-08
PRC1	2,74	1,12E-08
TK1	1,82	1,22E-08
CKAP2	3,08	1,58E-08
GTSE1	1,27	1,96E-08
WDHD1	1,05	2,16E-08
MCM6	2,02	2,17E-08
TACC3	1,54	2,31E-08
RPA2	2,35	2,9E-08
KIF18B	1,49	3,87E-08
CENPW	1,14	4,68E-08
TRIB3	1,18	6,05E-08
KIF11	1,44	6,24E-08
ASF1B	1,19	6,8E-08
MDM2	1,54	7,93E-08
KIF20B	1,48	8,59E-08
ARHGAP11B	2,1	1,05E-07
ZNF367	2,58	1,08E-07
PRIM1	2,35	1,37E-07
HLA-DRA	-1,98	1,81E-07
RAD18	1,28	1,92E-07
CENPK	1,9	2,06E-07