



Network and Pathway Analysis of Toxicogenomics Data

Gal Barel and Ralf Herwig*

Department Computational Molecular Biology, Max Planck Institute for Molecular Genetics, Berlin, Germany

Toxicogenomics is the study of the molecular effects of chemical, biological and physical agents in biological systems, with the aim of elucidating toxicological mechanisms, building predictive models and improving diagnostics. The vast majority of toxicogenomics data has been generated at the transcriptome level, including RNA-seq and microarrays, and large quantities of drug-treatment data have been made publicly available through databases and repositories. Besides the identification of differentially expressed genes (DEGs) from case-control studies or drug treatment time series studies, bioinformatics methods have emerged that infer gene expression data at the molecular network and pathway level in order to reveal mechanistic information. In this work we describe different resources and tools that have been developed by us and others that relate gene expression measurements with known pathway information such as over-representation and gene set enrichment analyses. Furthermore, we highlight approaches that integrate gene expression data with molecular interaction networks in order to derive network modules related to drug toxicity. We describe the two main parts of the approach, i.e., the construction of a suitable molecular interaction network as well as the conduction of network propagation of the experimental data through the interaction network. In all cases we apply methods and tools to publicly available rat *in vivo* data on anthracyclines, an important class of anti-cancer drugs that are known to induce severe cardiotoxicity in patients. We report the results and functional implications achieved for four anthracyclines (doxorubicin, epirubicin, idarubicin, and daunorubicin) and compare the information content inherent in the different computational approaches.

Keywords: network analysis, protein-protein interaction network, pathways, drug toxicity, toxicogenomics, transcriptomics, anthracyclines

INTRODUCTION

To thoroughly study the mechanisms behind drug induced toxicity a robust analysis by means of computational methods is crucial (Liebler and Guengerich, 2005). Understanding the influence of the compounds on different biological processes is complex and requires sophisticated interpretations of the data. In the field of toxicogenomics transcriptome data, that were collected upon drug treatment and that reflect gene expression levels in response to it, is in the focus of the analysis. Various studies, both *in vitro* and *in vivo*, focusing on different compounds and organs, have been already carried out (Hartung, 2009). Most of the studies were based on microarray technology (Mei et al., 2010), even though newer technologies, such as high-throughput sequencing

OPEN ACCESS

Edited by:

Danyel Jennen,
Maastricht University, Netherlands

Reviewed by:

Mohamed Diwan M.
AbdullHameed,
Independent Researcher, Frederick,
MD, United States

Richard John Brennan,
Sanofi, United States

*Correspondence:

Ralf Herwig
herwig@molgen.mpg.de

Specialty section:

This article was submitted to
Toxicogenomics,
a section of the journal
Frontiers in Genetics

Received: 19 June 2018

Accepted: 28 September 2018

Published: 22 October 2018

Citation:

Barel G and Herwig R (2018)
Network and Pathway Analysis
of Toxicogenomics Data.
Front. Genet. 9:484.
doi: 10.3389/fgene.2018.00484

(RNA-seq), are already in use in other research areas. Such transcriptomic profiles have previously been used for predicting toxic drug effects (Gusenleitner et al., 2014; Kohonen and Parkkinen, 2017; Nystrom-Persson et al., 2017; Rueda-Zarate et al., 2017), but further analysis for identifying the functional and molecular mechanisms behind the toxic effects is still much needed.

Here, we describe different approaches for the analysis of toxicogenomics data at the molecular network and pathway levels. We use publicly available data from microarray experiments and perform a differential expression analysis in order to identify the genes that are up- or down-regulated due to the administration of the drug (DEGs: differentially expressed genes). Suggested methods and tools include (i) over-representation analysis, (ii) gene set enrichment (pathway) analysis, and (iii) network propagation. All methods are complementary and deliver different and complementary mechanistic views on drug action and drug effects derived from the underlying gene expression data. Over-representation analysis provides a first impression on which pathways and biological functions are involved in the cell's response to the drug. This kind of analysis is typically done with statistical tests that evaluate the list of DEGs interrogating pre-defined gene sets that represent pathways or Gene Ontology (GO) functions. Gene set enrichment (pathway) analysis is a complementary approach in the sense that not only DEGs are investigated but rather the entire gene expression response. This ensures that not only those pathways appear interesting that agglomerate many DEGs but also those that agglomerate subtle but consistent transcriptome changes of many of their members (not necessarily DEGs). A third, more unsupervised approach is network propagation, a mathematical concept that traces the effects of perturbations (e.g., gene expression changes) simultaneously across a molecular network according to a specified rule. It assumes that a perturbation in a certain gene is not only affecting that particular gene but rather its entire network neighborhood. The signal induced by a perturbation is then propagated to the neighbors and the neighbors of the neighbors until a steady-state (or convergence state) is achieved. The result of the network propagation is a final state in which each node is assigned a final weight which can be used to identify highly affected nodes as well as specific interconnected parts of the network (network modules) that are mostly affected by the induced perturbations.

A previous effort to infer functional effects of drug treatments from gene expression data was done on a pathway level in the work of Hardt et al. (2016). They assembled the ToxDdb database, which contains gene expression data for more than 400 drugs and 2000 pathway concepts. This includes the association of drugs with specific molecular pathways, which can indicate to which mechanisms of action lead to toxicity. Here, we make use of this resource and the proposed implementation to extract pathways that are relevant for the toxic effect of different drugs. Additionally, we apply the same scoring scheme for measuring gene and pathway responses from gene expression data and enact a network analysis in order to identify functional modules that can also be associated with the toxic effect (see Methods).

Biological interactions are often described using molecular interaction networks (Barabasi and Oltvai, 2004), where each node represents a biological player, i.e., gene or protein, and each edge describes an interaction between a pair of nodes. Analyzing these networks can help to better elucidate the functional mechanisms that are being studied (McGillivray et al., 2018). There are numerous types of biological interaction networks (Vidal et al., 2011), as they can be based on different types of interactions and represent various biological actions. They vary between depicting gene regulatory interactions, viral-host interactions, metabolic reactions, protein-protein interactions, and more. Many of these interactions have already been made publically available through various specific databases, such as Reactome (Matthews et al., 2009), PID (Schaefer et al., 2009), KEGG (Kanehisa et al., 2012), and many others. Furthermore, there have been several attempts to combine and integrate different resources into one meta-resource, such as the work by Martha et al. (2011), IntNetDB (Xia et al., 2006), and ConsensusPathDB. In this work we make use of ConsensusPathDB (Kamburov et al., 2009), which currently integrates more than 600,000 interactions of different types which are collected from 32 public resources (Kamburov et al., 2013). Furthermore, we restrict our analysis to protein-protein interactions which are generally based on various experimental technologies (Walhout and Vidal, 2001).

One possible use of biological networks is for the identification of smaller subnetworks (subgraphs within the network), also referred to as modules, which depict an area that is more relevant for a specific biological function (Gustafsson et al., 2014). By integrating experimental data with interaction networks we can compute subnetworks that better represent the biological mechanisms which lead to a specific phenotype. There are several existing algorithms for module detection in biological networks. For a comprehensive overview of the different methods, see the recent review by Cowen et al. (2017). Many of these algorithms are based on a random walk process, where the weights of the nodes are propagated through the network, until a steady state is reached. The weighing of the nodes is dependent on the specific context and can be extracted for example from gene expression values or genetic mutation data. In this work we make use of the HotNet2 algorithm (Leiserson et al., 2015) that was originally developed for identifying subnetworks that result from somatic mutations. We apply the algorithm to toxicogenomics data and identify the most significant subnetworks for a drug treatment based on the gene expression response scoring.

We exemplify our approach on anthracycline drugs. Anthracyclines are a family of drugs that induce cardiotoxicity upon cancer treatment, and their use can result in cardiomyopathy and heart failure in many cases after a long period of time after treatment (Geisberg and Sawyer, 2010). These compounds are vastly used as chemotherapy agents, and have been shown to be extremely effective, but also to cause a major morbidity in cancer patients due to their toxic effects (Lenneman and Sawyer, 2016). Every exposure to anthracyclines carries some risk of resulting in cardiac dysfunction. The symptoms could present early on as well as at later times, in up to 23% of the patients (Steinherz et al., 1991). Although it is known that

anthracyclines disrupt the synthesis of DNA and RNA, mainly by inhibiting topoisomerase II (Geisberg and Sawyer, 2010) and that they lead to mitochondrial dysfunction (McGowan et al., 2017), the mechanisms that cause the cardiotoxic effects still remain largely unclear (Truong et al., 2014). Previous studies have tried to elucidate this problem, however, there is still need for further investigation so that detection and prevention could be improved (Raschi et al., 2010). We focus our analysis on the four most widely used compounds: daunorubicin, doxorubicin, epirubicin, and idarubicin. In addition, we compare the results to other chemotherapy agents from different drug families, which are also known to cause cardiotoxic effects.

MATERIALS

DrugMatrix

The toxicogenomics DrugMatrix (Ganter et al., 2005) database includes gene expression experiments from different rat tissue types at different time points and drug dosages. Data were downloaded via the diXa data collection (Hendrickx et al., 2015) that is available at <http://wwwdev.ebi.ac.uk/fg/dixa/index.html>. This data collection includes toxicogenomics profiles for 372 different compounds that were collected using the Affymetrix whole genome 230 2.0 rat GeneChip array. Data are available for heart, kidney, liver and muscle tissues, as well as for hepatocytes. The experiments were conducted for up to five times (after 0.25, 1, 3, 5 and 7 days) with only one dose concentration. In some cases, more than one dose was tested, and in others only one or two time points were measured.

Anthracycline Expression Data

The analysis was focused on four different anthracyclines compounds: daunorubicin (CHEMBL178), doxorubicin (CHEMBL53463), idarubicin (CHEMBL1117), and 4-epidoxorubicin (CHEMBL1237042). We downloaded the CEL files from the DrugMatrix database via the diXa data collection for these compounds in heart tissue. A full description of the treatments is given by **Table 1**. Daunorubicin is the only compound for which data are available at two doses, a higher “toxic” dose and a lower “pharmacological” dose (Ganter et al., 2005). Thus, for the analysis of daunorubicin we used only the higher dose.

Expression Data From Other Cardiotoxic Drugs

In order to evaluate the molecular effects that were identified in this work for the anthracyclines we also applied our workflow

to three other chemotherapy agents that are known to induce cardiotoxicities (Truong et al., 2014). Out of 41 drugs that are mentioned in the review by Truong et al., only these three had data available in the DrugMatrix database. Cyclophosphamide (CHEMBL88) and ifosfamide (CHEMBL1024) are both alkylating agents, and imatinib (CHEMBL941) is from a family of small-molecule targeted therapy drugs. We downloaded the CEL files from the DrugMatrix database for these compounds in heart tissue. A full description of the treatments is given by **Table 2**. Data for imatinib were available in two different doses, and so we applied our analysis for the higher dose only.

ConsensusPathDB – A Molecular Interaction Network Resource

ConsensusPathDB (Kamburov et al., 2009) is a meta-database for molecular interactions and pathways that currently integrates 32 public resources (Kamburov et al., 2013) and is composed of more than 600,000 unique interactions of different types and holds more than 5,000 human pathway concepts. The database is available through a web server¹ where queries of genes, proteins, drugs and other types of biomolecules can be made, along with gene and metabolites analysis, such as enrichment and over-representation analysis (Herwig et al., 2016).

ConsensusPathDB holds an integrated network which is comprised of more than 300,000 binary protein–protein interactions (PPIs) representing a comprehensive model of the human interactome. These interactions were scored with a mixture of topology-based and annotation-based measures, such as the ones described in Goldberg and Roth (2003), Kuchaiev et al. (2009), Yu et al. (2010), and Kamburov et al. (2012a). These measures were aggregated into a meta-score using the IntScore (Kamburov et al., 2012b) approach, which combines the individual confidence scores, and provides a final score that better indicates how plausible the interaction is. The PPI network, along with the quality assessment scores, can be downloaded via http://cpdb.molgen.mpg.de/download/ConsensusPathDB_human_PPI.gz.

ToxDB

ToxDB (Hardt et al., 2016) integrates toxicogenomics data from two large-scale studies, Open TG-GATEs (Uehara et al., 2010) and DrugMatrix (Ganter et al., 2005), with pathway concepts from ConsensusPathDB (Kamburov et al., 2009). It contains a total of 7,464 different treatment data sets, covering 437 drugs, and 2,694 molecular pathway concepts with response scores. Its web interface is available at <http://toxdb.molgen.mpg.de/and>

¹<http://consensuspathdb.org>

TABLE 1 | Anthracyclines drug treatment experiments from the DrugMatrix database.

Drug	Time points (days)	Dosage (mg/kg)
Daunorubicin	1, 3, and 5	2/3.25
Doxorubicin	1, 3, and 5	3
Idarubicin	1, 3, and 5	0.625
4-Epidoxorubicin	1, 3, and 5	2.7

TABLE 2 | Other chemotherapy drugs and their experiments information from the DrugMatrix database.

Drug	Time points (days)	Dosage (mg/kg)
Cyclophosphamide	3 and 5	25
Ifosfamide	3 and 5	143
Imatinib	1, 3, and 5	15/150

allows browsing for the effect of a drug treatment on cellular pathway response. The user can also browse for a specific pathway and retrieve the treatments that affect it the most.

METHODS

Microarray Data Processing

We processed the microarray data sets of the heart tissues that were treated with anthracyclines. The oligonucleotide sequences (oligoprobes) that were downloaded from DrugMatrix were mapped to the rat genome-build and probe sets were redefined using the resource at <http://brainarray.mbni.med.umich.edu/CustomCDF> such that each probe is assigned to a unique gene, and each gene is associated with a varying number of probes. It has been shown that re-mapping of oligoprobes unambiguously to the latest genome-build increases performance of Affymetrix Gene Chip transcriptomics platforms (Dai et al., 2005). The replicates of the different drug treatments were grouped together, according to their corresponding dosage and time point. The raw data were normalized using the GC Robust Multi-Array method in the R package *gcrma* (Gentry, 2017).

Orthology Mapping

Rat genes had to be mapped to human genes by orthology, in order to use the human pathway concepts and PPIs from ConsensusPathDB. This was done via the orthology mapping of the Ensembl BioMart repository (Yates and Akanni, 2016). We used only “One2one” and “One2many” homology relationships: if the rat gene has exactly one orthologous human gene, the corresponding rat microarray value is assigned to that human gene. Otherwise, if the rat gene has multiple orthologs in the human genome, the corresponding rat microarray value is assigned to all human paralogs.

Differential Gene Expression Analysis

The normalized microarray data were analyzed with the R package *limma* (Ritchie et al., 2015) in order to calculate differentially expressed genes (DEGs), i.e., genes that are up- or downregulated significantly when comparing compound treatment against control experiments. It estimates fold-changes and standard errors by fitting a linear model to each gene profile and uses an empirical Bayesian approach to smoothen these errors.

We applied *limma* for every pair of case-control normalized microarray values. Therefore, for every gene, given any drug, dosage and time point combination, we can calculate its fold change value and a corresponding *P*-value. Fold change is computed as the ratio of the mean expression values of treatment and control. *P*-value is the significance of the fold change given the null hypothesis that there is no change in expression between treatment and control.

Gene Scoring

In order to measure the response of a gene to a drug treatment experiment we use the following scoring scheme:

for every gene *i*, every drug *j* and every time-dosage treatment *k*:

$$S_{ijk} = |\log_2 r_{ijk}| |\log_{10} P_{ijk}| \quad (1)$$

Here r_{ijk} is the fold change between the treatment and the control experiments, and P_{ijk} is the *P*-value from the differential expression analysis. This score describes a weighted fold change of the gene, such that the more significant the change is, the higher the weight is. Using this scoring scheme allows us taking into consideration the rather low sample size of the experiments, as well as to avoid a pre-selection of the genes based on their *P*-values only. The score serves as a measure of how much the gene was affected by the treatment, regardless of the change in expression (higher or lower expressed in comparison to the control).

Pathway Scoring

In previous works (Yildirimman et al., 2011; Hardt et al., 2016) we have also defined a pathway scoring scheme, which is based on the scoring of the genes that the pathway is comprised of. Here, we take all available human pathways from ConsensusPathDB, and their associated genes. We compute for each pathway a relative pathway response (RPR) score which serves as a measure for the response of the pathway to the drug, given gene expression data. The higher the RPR score is, the more significant is the response of the pathway to the treatment. A pathway M_1 is defined as a set of *m* genes: $M_1 = \{g_1, \dots, g_m\}$. Given a treatment of drug *j* at a time point and dosage *k*, we can calculate the pathway score:

$$M_{1,j,k} = \frac{1}{m} \sum_{g_i \in M} S_{ijk} \quad (2)$$

Where S_{ijk} is the gene score of gene *i*, as defined in Equation 1. The RPR score of the pathway M_1 with respect to the drug *j* and the time-dosage *k* is calculated by dividing the pathway score by $\bar{M}_{j,k}$ the median of all pathway scores, given drug *j* and time-dosage *k*:

$$RPR_{1,j,k} = \log_2 \left(\frac{\bar{M}_{1,j,k}}{\bar{M}_{j,k}} \right) \quad (3)$$

In addition, we computed RPR scores for all pathways in all the different experimental conditions and derived a background distribution. This background distribution is used to judge the significance of a given RPR score and reflects the response of the pathway to the experimental condition.

Network Module Analysis

A network module analysis was carried out by applying the HotNet2 (Leiserson et al., 2015) algorithm, which was originally developed to identify significantly mutated subnetworks in cancer in PPI networks based on somatic mutations data. The algorithm takes as input a score vector $S = (S_1, \dots, S_n)$, where *n* is the number of genes, and a graph $G = (V, E)$. The gene scores are computed context dependent (see below), and the graph represents a PPI network, where each node corresponds to a protein coding gene, and each edge to an interaction between

their respective proteins. HotNet2 then applies an insulated heat diffusion process that includes the following steps:

1. Heat diffusion – at each time step heat is diffused from every node i to every one of its neighbors j . The amount of heat that will be placed on node j given the initial heat on node i is given by the entry (i,j) of the diffusion matrix F , which is defined by:

$$F = \beta (I - (1 - \beta) W)^{-1} \quad (4)$$

$W_{i,j} = \frac{1}{deg(j)}$ if (i,j) are neighbors, otherwise 0.

The parameter β is an insulating parameter and W is the normalized adjacency matrix of the input graph G such that $deg(j)$ is the degree (number of neighbors) of node j . I is the identity matrix.

2. Exchanged heat – the amount of heat that diffuses from node j to node i when heat S_j is placed on node j is given by the exchanged heat matrix E which is defined by:

$$E = FD_s \quad (5)$$

Where D_s is a diagonal matrix with the entries of S .

3. Identification of subnetworks – a new weighted directed graph H is created using the nodes V . Node i will be connected to node j in this graph if $E(i,j) > \delta$, where δ is a minimum edge weight parameter, and their respective edge will have a weight equal to $E(i,j)$. Then, the strongly connected components of H are identified and are selected to be the final subnetworks.

4. Statistical test for the subnetworks – a two-stage statistical test, that is described in the original HotNet algorithm (Vandin et al., 2011, 2012), is applied to determine the significance of the number and the sizes of the subnetworks.

To identify functional modules that are associated with the different drug treatments we used the HotNet2 algorithm (Leiserson et al., 2015) that is available at <http://compbio.cs.brown.edu/projects/hotnet2/>. Since the first step of the algorithm depends only on the graph G and the chosen parameter β , we calculated the diffusion matrix F for the high-confidence ConsensusPathDB PPI network, while choosing $\beta = 0.5$. For the scoring of the genes, we used our own data-derived scores: for each drug and treatment, we used as input their gene scores, as described in Equation 1. The output of the HotNet2 algorithm depends largely on δ , the minimum edge weight parameter. The lower its value, the larger are the subnetworks. HotNet2 outputs four different results, for four different δ values, which are chosen based on a permutation test in their algorithm [for further details see (Leiserson et al., 2015)]. In this work we chose for further analysis the subnetworks which are resulted when taking the smallest δ parameter from the output of the HotNet2 algorithm.

Over-Representation Analysis (ORA)

ConsensusPathDB allows performing over-representation analysis (ORA) with different functionally relevant gene sets (Herwig et al., 2016). Given a set of genes, proteins or metabolites over-represented sets are searched among three pre-defined categories: (1) network neighborhood-based sets, (2) pathway-based sets, and (3) Gene Ontology (GO)-based sets. According to the hypergeometric test, a P -value is calculated based on the number of identifiers that are present in the given set and in the

pre-defined sets. As background, the user can choose another set of identifiers, for example all genes that were measured in the experiment, or simply use all entities that are annotated in ConsensusPathDB. In our work, ORA was used to identify only the pathway based enriched sets, choosing all possible pathways from ConsensusPathDB and applying a P -value cutoff of 0.01. As background, we used the full list of genes that were measured in the corresponding experiment.

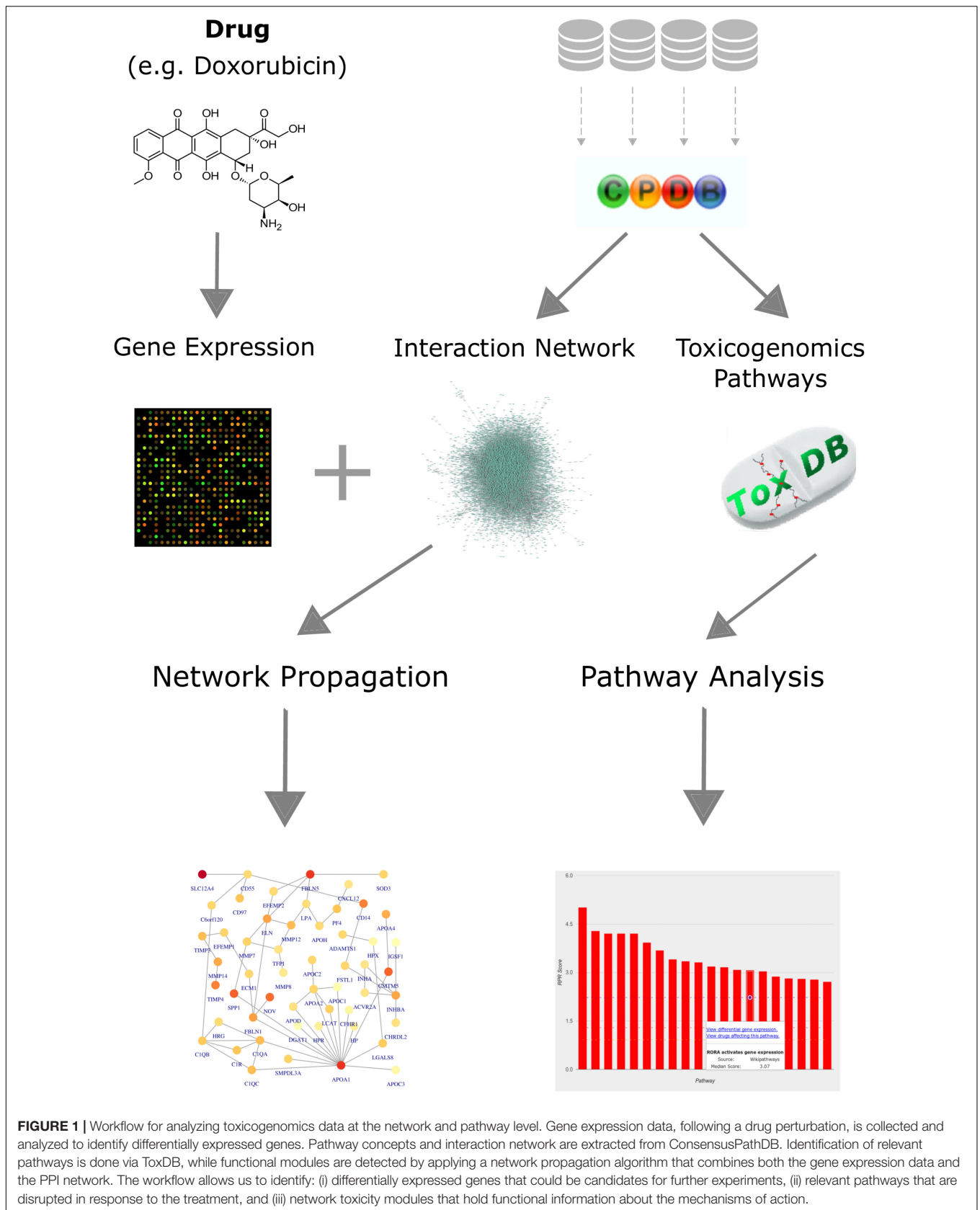
RESULTS

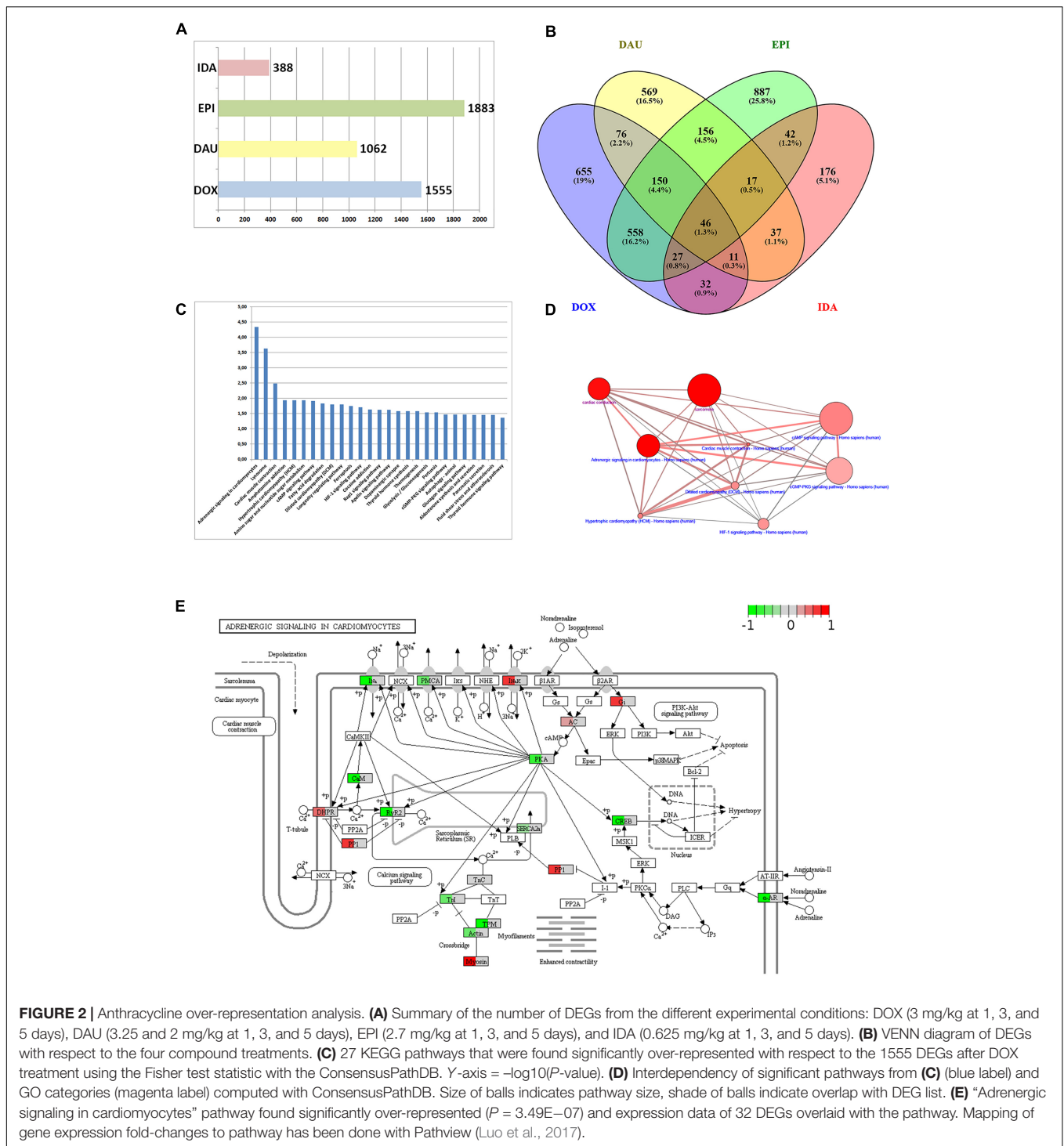
Workflow for Analyzing Toxicogenomics Data in the Context of Networks and Pathways

We established a computational workflow for analyzing toxicogenomics data by incorporating pathway and network information using different complementary approaches in order to gain functional information from gene expression data (**Figure 1**). We exemplify the results on the four anthracycline drugs: daunorubicin (DAU), doxorubicin (DOX), idarubicin (IDA), 4-epidoxorubicin (EPI). We also applied our analysis to three other anti-cancer drugs that are known to cause cardiotoxicity: cyclophosphamide (CYC), ifosfamide (IFO), and imatinib (IMA). We compare our results for the anthracyclines with our results for these drugs in order to identify differences and commonalities and distinguish the effects that are explicit to anthracyclines. The workflow is based on the results of a differential expression analysis, and combines pathway and network information from both ConsensusPathDB and ToxDB. It begins with an over-representation analysis for the DEGs, using pathway concepts that are collected in ConsensusPathDB, in order to assign a biological function to the most significantly changed genes. Next, it continues with a pathway analysis using ToxDB, extrapolating from DEGs to the entire gene expression response and from gene lists to pathway concepts. Using molecular interaction information from ConsensusPathDB, the workflow also includes a PPI network construction and an analysis that applies a network propagation algorithm which combines the DEGs with the PPI network. Finally, it is able to identify subnetworks that we define as drug toxicity modules.

Assigning Biological Function to Gene Lists With Over-Representation Analysis

A first step in functional interpretation of toxicogenomics results is to interrogate the lists of DEGs (see Methods) for known annotation sets such as pathways or GO terms using Fisher's test or similar statistics (see Methods). Summarizing the different experiments (time points and dosages) for the four anthracyclines (DOX, DAU, EPI and IDA) results in 1,883 DEGs for EPI, 1,555 for DOX and 1,062 for DAU whereas IDA shows a much weaker response with 388 genes (**Figure 2A**). In all cases, human genes were inferred based on homology mapping of the corresponding rat microarray probes. All anthracyclines were administered at maximum tolerated doses (MTDs) and, thus should be of comparable toxicity (DOX 3 mg/kg; DAU 3.25 mg/kg; EPI





2.7 mg/kg and IDA 0.625 mg/kg). Also, it has been shown that gene expression signatures are predictive of toxicity and that number of DEGs is indicative of phenotypically observed injury of the organ (Paules, 2003; Andersen et al., 2008; Holmgren et al., 2015) which has given rise to the concept of phenotypic anchoring, i.e., the association of gene expression signatures to toxic phenotypes. The difference in DEGs between DOX and IDA

is in line with previous findings: For example, Platel et al. (1999) showed that in rat the MTDs for DOX and IDA were 3 mg/kg and 0.75 mg/kg, i.e., comparable to the levels used in the DrugMatrix screen, and that at these MTDs IDA showed significantly lower cardiotoxicity than DOX. Anthracyclines show highly specific response at the gene expression level (**Figure 2B**) with 40–50% of all DEGs specific for a certain compound. The strongest relative

agreement in gene expression response was observed between DOX and EPI (45%), whereas the relative agreement between IDA and the other three compounds is much lower (12–15%). DOX response can be characterized at the pathway level using ORA analysis (see Methods). **Figure 2C** exemplifies the results for DOX using gene sets that represent KEGG (Kanehisa et al., 2012) pathways and that might reflect the cell's response to the drug treatment. In total, 27 KEGG pathways are significantly over-represented (Q -value < 0.05 and at least 10 DEGs overlapping with the pathway gene set). A number of disease gene sets have been identified such as “Hypertrophic cardiomyopathy” ($Q = 0.0115$), “Dilated cardiomyopathy” ($Q = 0.0158$) or “Cardiac muscle contraction” ($Q = 0.0032$). Interestingly, these pathways were also found in a recent study investigating cardio-toxicity in human pluripotent stem cell derived-cardiomyocytes (Maillet et al., 2016) and thus seem to extrapolate from rat *in vivo* to human *in vitro* studies. The top-enriched pathway in our setting is “Adrenergic signaling in cardiomyocytes” ($Q = 4.54E-05$). 32 genes of that pathway are differentially expressed including troponins (*TNNC1* and *TNNI3*), tropomyosins (*TPM1* and *TPM2*), and other well-known toxicity-associated genes such as *RYR2* (ryanodine receptor 2). **Figure 2D** displays the interdependencies of these and other disease-related gene sets.

An important feature in the analysis is the visualization of the gene expression changes in the pathway map. Pathway maps can be retrieved by pathway resources such as KEGG. There are many tools that allow visualizing gene expression fold-changes on these pathways which is exemplified in **Figure 2E** with the “Adrenergic signaling in cardiomyocytes” and the expression fold changes of the DOX treatment.

From Genes to Pathways – Pathway Analysis Using the ToxDB

The next level of analysis is to extrapolate the gene expression values from single genes to entire pathways. We have built a tool, ToxDB that combines gene expression data and pathway concepts. ToxDB builds on three components: (i) a comprehensive collection of pathway concepts along with drug treatment microarray data, (ii) a numerical method to compute pathway responses from genome-scale expression data, and (iii) a web interface that allows user interaction. By this procedure each pathway is assigned a numerical value that reflects its response to the treatment (see Methods). ToxDB contains pre-calculated pathway scores for ca. 2,700 different pathways and ca. 7,500 experimental conditions mainly extracted from two large toxicogenomics studies, TG-GATES and DrugMatrix (see Methods). A background distribution of pathway scores is used to infer statistical significance. ToxDB can be used in different views. The drug view allows drug centric analysis: by selecting a compound, for example DOX, and a specific experiment all responding pathways can be viewed (**Figure 3A**). By further clicking on a specific response pathway [here “Hypertrophic cardiomyopathy (HCM)”] the expression results of all genes can be inspected that are associated with this pathway (**Figure 3B**). DEGs of this pathway are known cardiac-relevant genes such as *MYH7* (myosin, heavy chain 7, cardiac muscle,

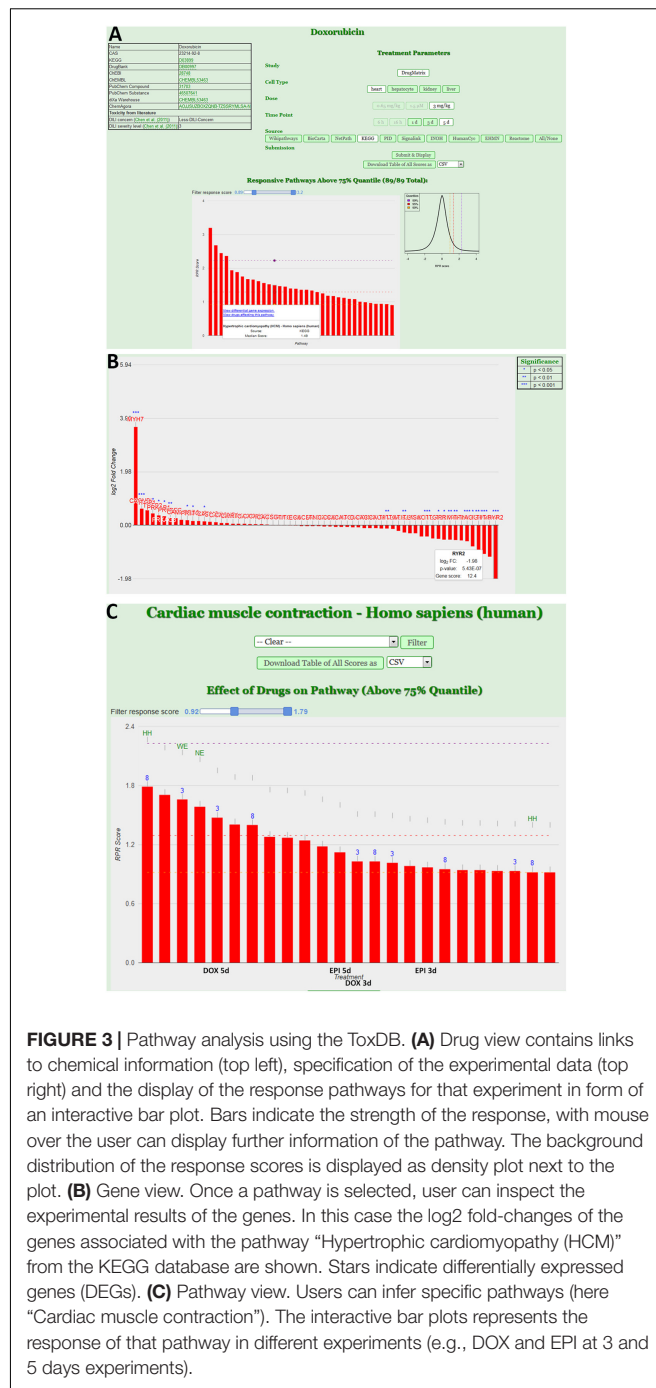


FIGURE 3 | Pathway analysis using the ToxDB. **(A)** Drug view contains links to chemical information (top left), specification of the experimental data (top right) and the display of the response pathways for that experiment in form of an interactive bar plot. Bars indicate the strength of the response, with mouse over the user can display further information of the pathway. The background distribution of the response scores is displayed as density plot next to the plot. **(B)** Gene view. Once a pathway is selected, user can inspect the experimental results of the genes. In this case the \log_2 fold-changes of the genes associated with the pathway “Hypertrophic cardiomyopathy (HCM)” from the KEGG database are shown. Stars indicate differentially expressed genes (DEGs). **(C)** Pathway view. Users can infer specific pathways (here “Cardiac muscle contraction”). The interactive bar plots represents the response of that pathway in different experiments (e.g., DOX and EPI at 3 and 5 days experiments).

beta; \log_2 -FC = 3.65, $P = 9.05E-06$), *DES* (desmin; \log_2 -FC = 0.264, $P = 7.55E-02$), *TPM4* (tropomyosin 4; \log_2 -FC = -1.07, $P = 9.14E-04$), or *RYR2* (ryanodine receptor 2; \log_2 -FC = -1.98, $P = 5.43E-07$). A second view is the pathway view: the user can select a single pathway (here “Cardiac muscle contraction”) and as a result all experiments are shown in which this pathway responded significantly (**Figure 3C**). Pathways can be selected from ten different resources which comprise most widely used pathway resources such as KEGG, Reactome

or BioCarta. It can be seen from the view that anthracycline experiments (DOX and EPI at different time points) are among the compounds that induce the most significant responses of cardiac muscle contraction.

Protein–Protein Interaction Network Construction

Protein–protein interaction networks are typically used as scaffolds for drawing network propagation of gene expression data on. The underlying argument is “guilt-by-association,” i.e., the assumption that genes/proteins that interact with each other usually share function and, thus, that network modules computed from these PPI networks amplify functional information. Thus, the PPI network needs to be properly selected in the sense that it should be sufficiently comprehensive and that the false-positive rate of interactions should be low.

In this work we make use of the PPI network from ConsensusPathDB (release 32) and reduce it to a high-confidence network by taking only the interactions with a confidence score of 0.95 or higher (Figure 4C). This network is comprised of 10,707 proteins and 114,516 unique interactions (Figure 4A). Biological networks are normally characterized with a power law distribution of the node degree (Barabasi and Oltvai, 2004). This means that most of the nodes in the network are only connected to a few other nodes, while a small majority is very highly connected, with more than 400 neighbors (Figure 4B). We make use of this high-confidence interaction network in our workflow in order to identify subnetworks that are highly relevant to the drug treatments.

Toxicity Network Modules Are Identified by Applying Network Propagation

Toxicity modules were calculated using the HotNet2 algorithm for each drug and time point independently. A detailed list of all the toxicity modules is provided in **Supplementary Table 1**. Here we discuss the results when using the gene expression values for DOX only. The modules for the other anthracyclines are provided in **Supplementary Figures 1–3**. Since the drug treatments of DOX were measured three times over the course of 5 days, we derived one module for each one of the time points (Figure 5A). Looking at each one of the modules, and at all of them together, allows us to analyze the changes over time. We identified that the effect becomes stronger after 3 days, as the size of the module grows, but also that it is again much lower after 5 days. This could be due to the toxic effect of the drugs on the cells, i.e., the cells might already be dying. We confirmed this by looking at the over-represented pathways for the genes in the “5 days” module (Figure 5B). We observed two pathways that indicate cell death: “Apoptosis” and “Apoptotic Signaling Pathway.” In addition, we identified another pathway that might be involved in cardiotoxicity: “Cardiac Progenitor Differentiation” (from the WikiPathways database). This pathway includes several factors that are involved in cardiac differentiation, such as *TNNI3* that we also detected as differentially expressed, and is based on two recent reviews (Burrige et al., 2012; Stillitano et al., 2012). The

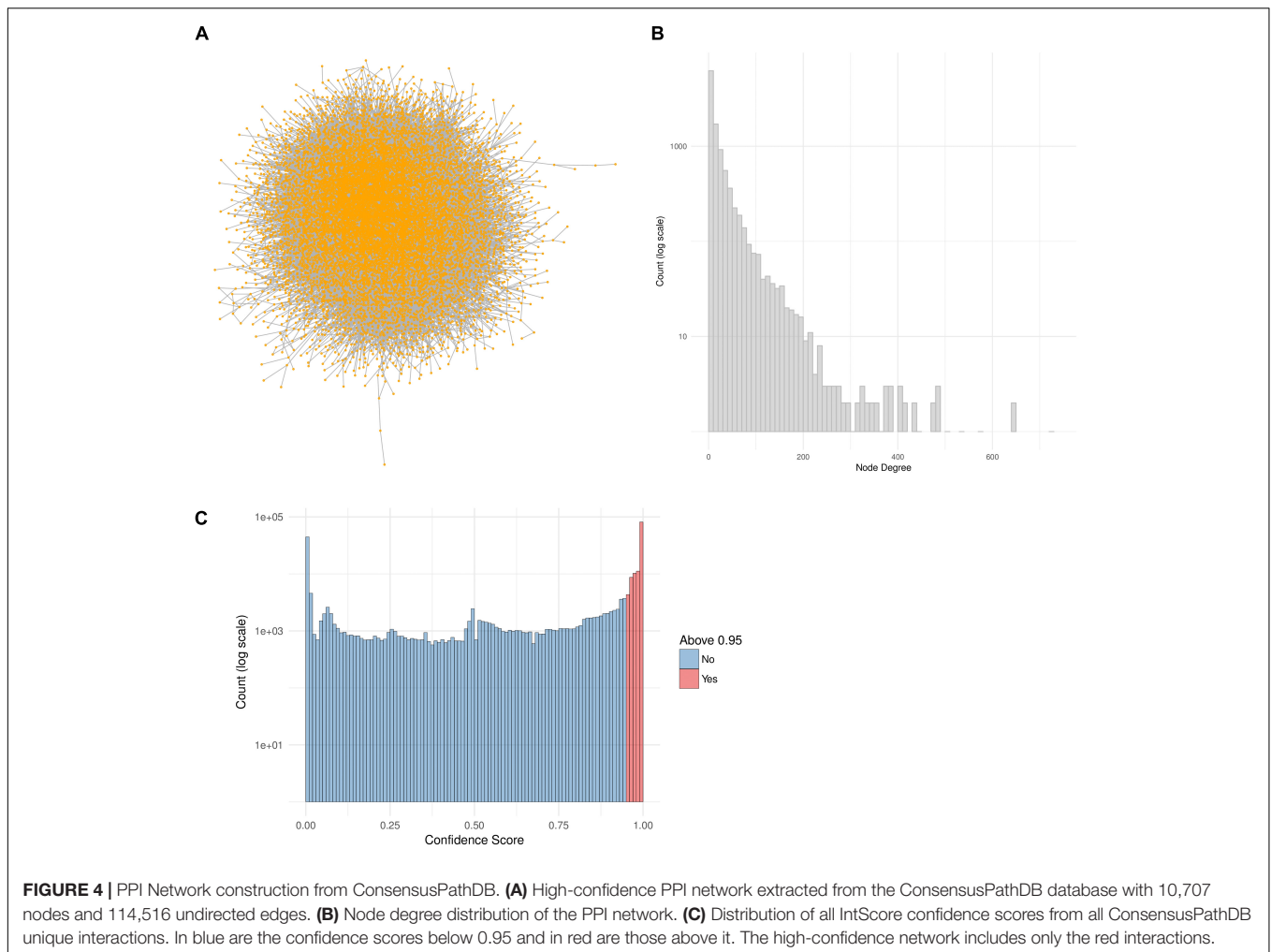
module also includes the genes *IGF1* and *IGF2* that are involved in the differentiation of immature cardiomyocytes and have been associated with cardiac hypertrophy (Wang et al., 2012). Other genes that might be involved in cardiotoxicity and are present in the “1 day” and “3 days” modules are *APOA1*, that have been previously associated with hereditary amyloid cardiomyopathy (Hamidi Asl et al., 1999), and *ELN* that has been involved in both progressive aortic valve malformation and latent valve disease in mice (Hinton et al., 2010).

Network Modules Amplify Functional Information

We compared the over-represented pathways when using only the high scoring genes (genes with a score above the 99th quantile of the background distribution of all scores), and when using the genes from the network modules (Figure 6). In 8 out of 12 drug-treatment conditions the enrichment scores when using the genes from the network modules, were higher than the scores when using the high scoring genes only. Furthermore, when comparing the significance of the enrichment, by looking at the means of the Q-values (FDR corrected *P*-values), in all but one case we observed a higher enrichment when using the genes from the modules. This suggests that the network modules are enriched in more functional information, and therefore they serve as a powerful mean for studying systemic processes, such as drug induced toxicity.

Differences and Commonalities Between Anthracyclines and Other Chemotherapeutic Drugs

To assess the specificity of our results for anthracycline-induced cardiotoxicity, we applied the same workflow (Figure 1) to three other anti-cancer drugs which are known to cause cardiotoxic phenotypes: cyclophosphamide, ifosfamide and imatinib (see Materials). Looking only at the high scoring genes (genes with a score above the 99th quantile), with gene scores computed according to section 3.4, we observe hardly any common genes between the three drugs (Figure 7A). Interestingly the 17 genes that are common between these three drugs are also common with all the other anthracyclines drugs. However, when we compare the genes that are present in the toxicity modules of these drugs and the toxicity modules of all anthracyclines (Figure 7B) we detect 214 common genes. This highlights the fact that the network propagation approach amplifies gene expression responses toward relevant cardiotoxic mechanisms and phenotypes that are shared by the different drugs so that different gene expression responses can result in similar pathway responses. Evidently the number of genes in the anthracyclines modules is much higher as they are derived from more drugs and experiments, but nonetheless the percentage of number of genes that are shared is much higher. This could again indicate to the functional information that is inherent within the toxicity modules, which might suggest to the mechanisms that are involved in causing the toxic effect.



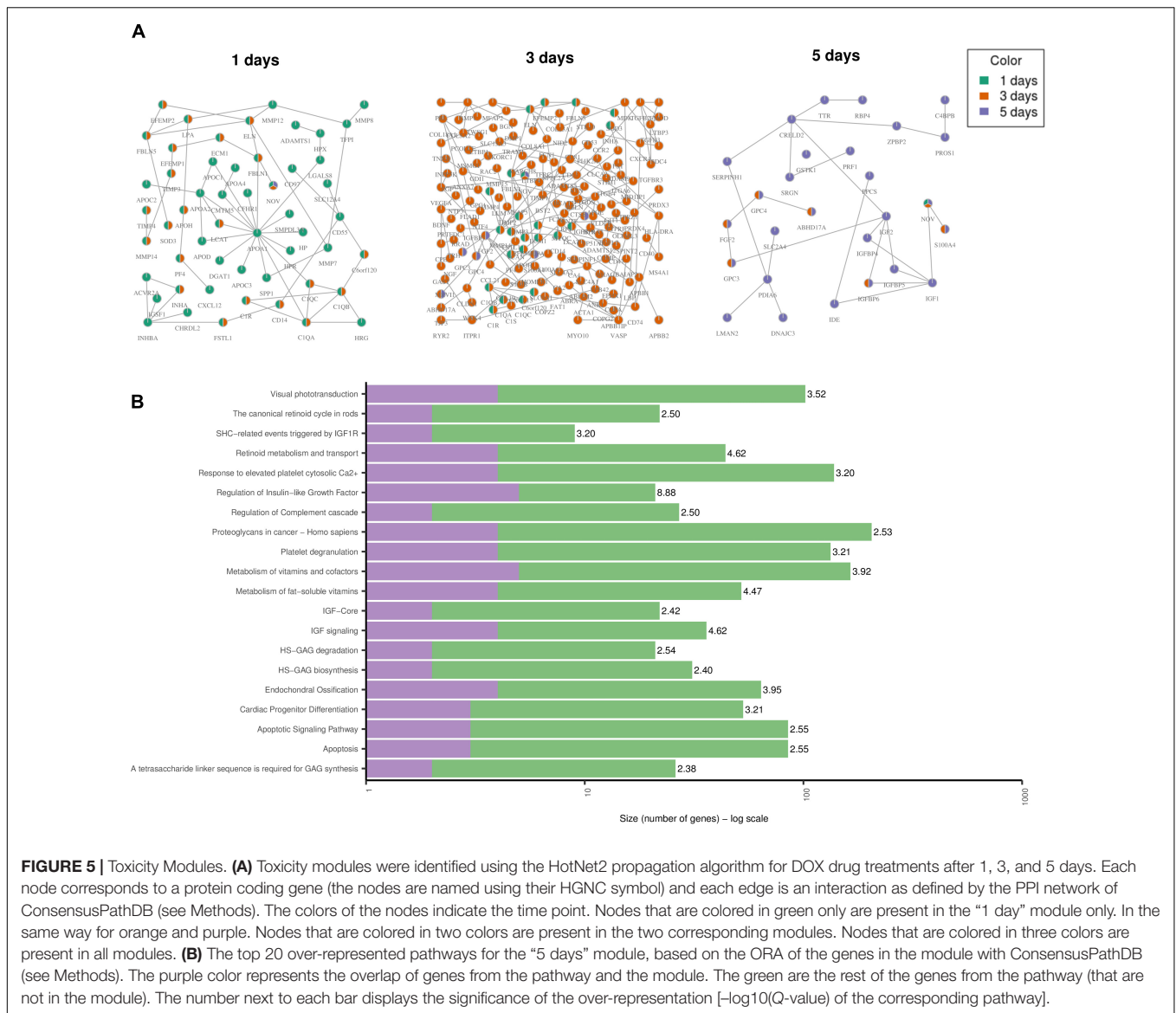
In order to identify biological functions that are specific for anthracyclines we performed enrichment analysis with the set of 330 genes that are solely part of the anthracycline toxicity modules (**Figure 7C**). We observe enrichment of cardiac disease pathways such as “Viral myocarditis,” “Hypertrophic cardiomyopathy,” and “Dilated cardiomyopathy,” mainly through the inclusion of ITGB and TGF β gene family members and RYR2. Another strong signal is the presence of immune response pathways. It is well-known that anthracycline treatment can induce systemic inflammation mediated through interleukins (Mills et al., 2008; Sauter et al., 2011). Interestingly, many inflammatory and immune response pathways are enriched with the anthracycline toxicity modules, in particular through interleukins (IL1A, IL12A, IL12B, IL23A, IL33, and IL27RA) that are not included in the modules of the other drugs.

DISCUSSION

Combining the information from ConsensusPathDB and ToxDB, including pathway concepts and a PPI network, together with

experimental data, allows for a more comprehensive view of the effects of the drug treatments. Firstly, by using ToxDB we are able to identify pathway concepts and by that suggest specific mechanisms that may be either the cause or the consequences of the toxic effects. In addition, by using the information from the PPI network and a propagation algorithm, we can also identify specific interactions that could be highly relevant for further experiments. These network modules carry out more functional information, since their genes and interactions represent parts of different pathways, and thus they are enriched in more information about specific biological mechanisms. Indeed, by propagating perturbation data across a network it is possible to gain information not only for the genes that were actually measured by the experiment but in addition also for the genes that haven’t been measured experimentally but that are connected with many measured neighbors in the network.

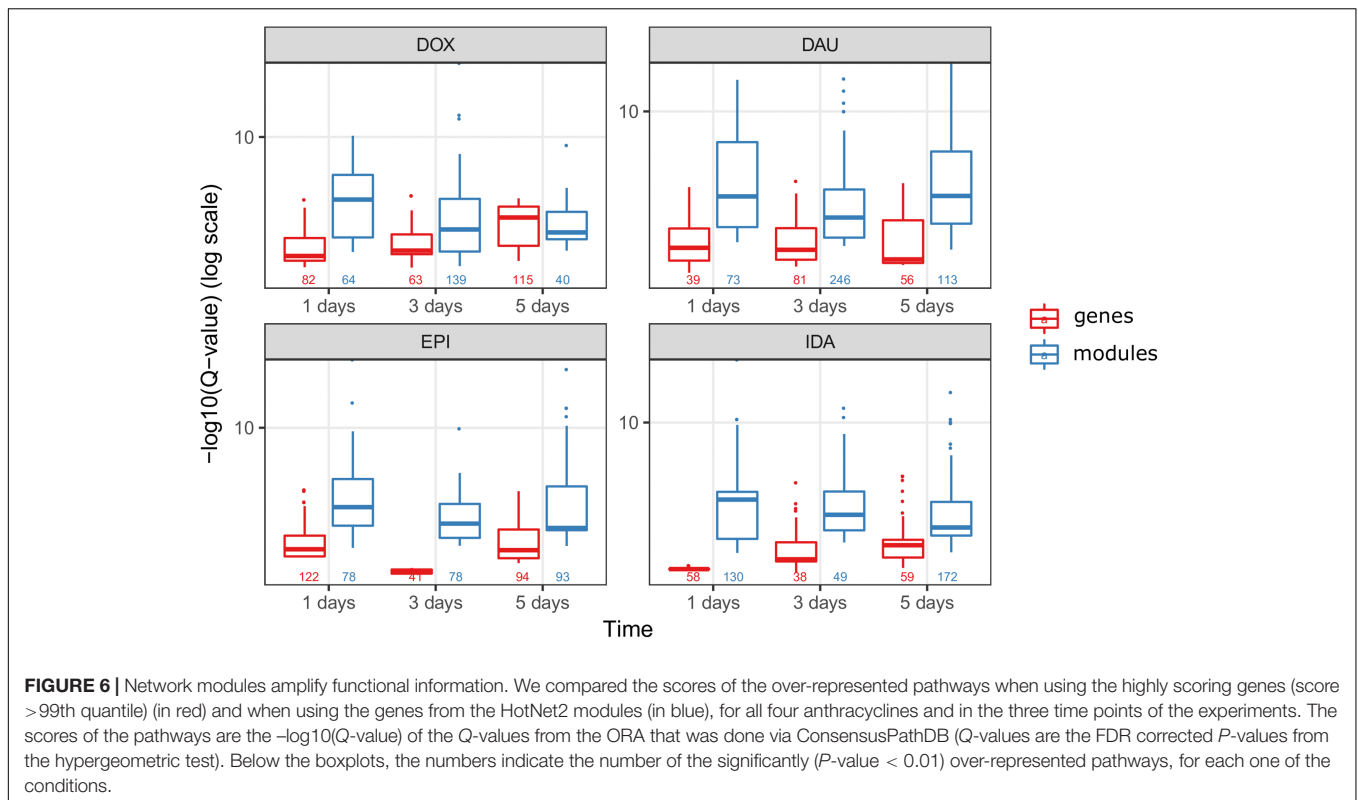
When looking only at DEGs, it is very difficult to describe the toxic effects of a drug given a specific treatment. Usually, this list of genes is comprised of hundreds of possible candidates, and it can be very challenging to distinguish which ones are involved in causing toxicity. Other works have tried to reduce the number of genes by looking at a smaller toxicogenomics



space (Kohonen and Parkkinen, 2017). By defining a more complex gene score, we were able to reduce the number of genes such that it becomes easier to extract plausible candidates for further studies. Furthermore, by applying a network propagation scheme to the gene scores and the high-confidence PPI network, we were able to both reduce the list even further, and also identify functional modules within PPI networks. These functional modules can better reflect the mechanisms that lead to toxicity, as they contain not only the obvious candidate genes based on the differential expression analysis, but also other genes that might be associated with the toxic effect, and are also connected to the more significantly changed genes.

ConsensusPathDB is a meta-database that agglomerates information from multiple resources and therefore includes different kinds of interactions: protein-protein, genetic, metabolic, signaling, gene regulatory and drug-target. In

addition, it also holds information about biochemical molecules and pathways. The high confidence PPI network that we have constructed is comprised solely of highly scored protein-protein interactions that are extracted from several resources, such as BIND, INTACT, HPRD. However, the ORA that is provided within ConsensusPathDB, searches for over representation of genes within cellular pathways. These pathways are derived from other resources, such as KEGG, Reactome, WikiPathways, etc. The different resources are completely independent data sets and the ConsensusPathDB simply serves as a common analysis platform. Therefore, when we apply the ORA to the extracted network modules, we can identify how enriched they are not only with protein-protein interactions, but also with pathway information. As we have illustrated in **Figure 6**, network modules contain not only protein-protein interaction information (that is inherent within its structure) but also are enriched in



other functional information that is represented in various pathways.

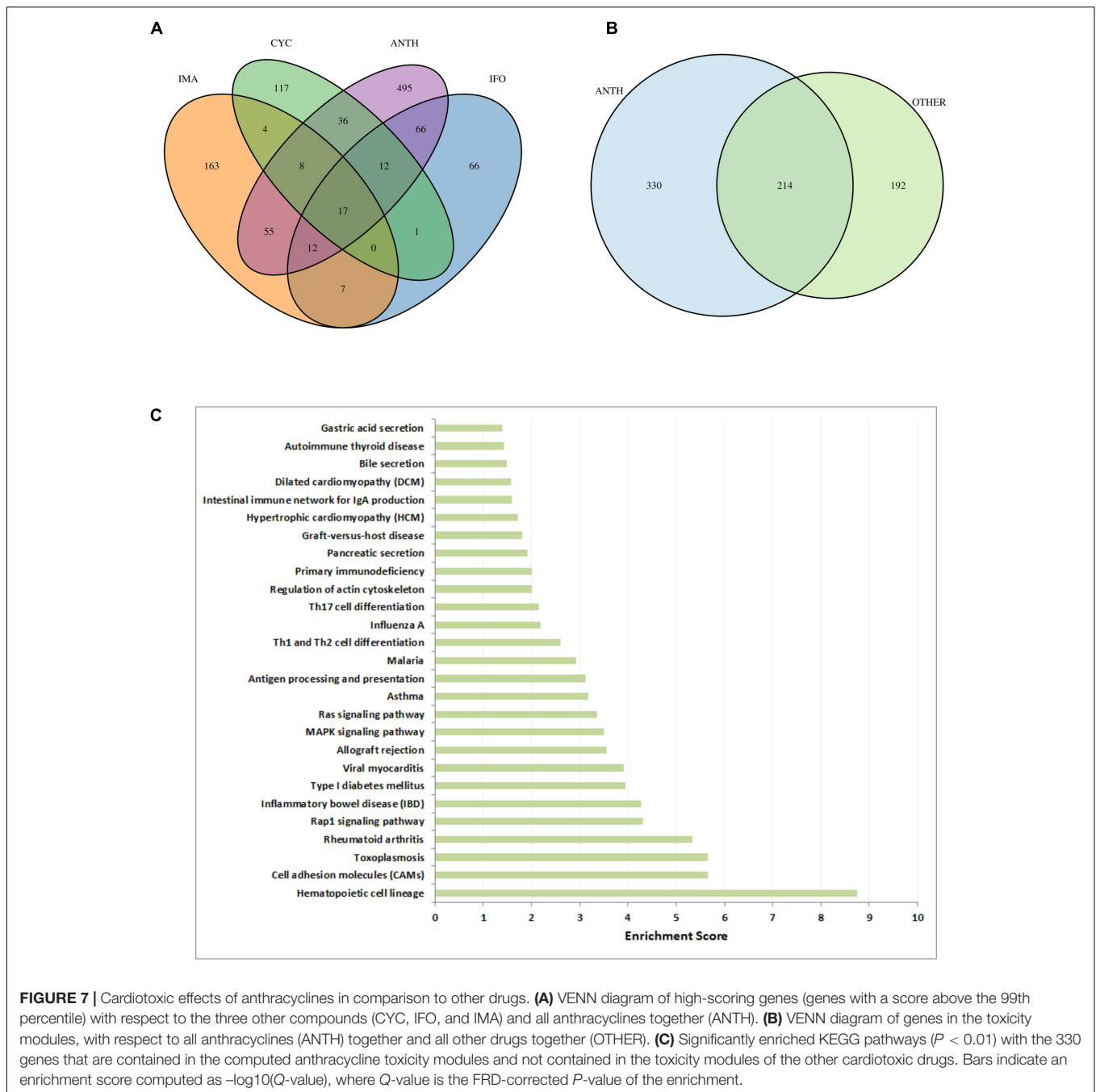
It should be noted that besides the described publicly available tools for pathway annotation and analysis, there are commercially available tools that hold functionality for pathway and network analysis such as IPA (Ingenuity/Qiagen), TransPath (geneXplain), or MetaCore (Thomson Reuters). These and other commercial and publicly available tools can be used to construct suitable molecular networks and perform enrichment analysis and module computation. A survey of databases and resources is given by Pathguide (Bader et al., 2006), a recent review and comparison of pathway tools has been published for example for metabolomics data (Marco-Ramell et al., 2018).

Toxicology studies often explore the effects of compounds over time and varying dosages (Hartung, 2009). Here, we analyzed gene expression levels for three different time points: after 1, 3, and 5 days. Every time point experiment was independently compared to the control experiment, such that a network module was constructed for every time point. To discern the effect over time, we compared between the modules, and determined the possible changes due to time. We were able to identify the toxic effect over time by looking at the different modules and also the genes within the module that could implicate the pathways that are leading to toxicity. In the future, one could try to first integrate the experimental data from the different time points, such that the change in expression levels over time is taken under consideration. For example by applying a mathematical model to detect differential

expression over time, like the one suggested by Conesa et al. (2006). We could further use the results of such model and incorporate them into the network propagation algorithm in order to identify a module that encompasses data from all the time points together.

The approach we applied in this work consists of three main components: gene expression analysis, a PPI network and a network propagation algorithm. All of these have several alternatives, and could be further incorporated in future analysis. Firstly, the PPI network can be replaced with other genetic interaction networks, for example a gene regulatory network that is derived from experimental data (Zheng and Huang, 2018). Secondly, different types of experimental data can be used for ranking the genes and using their ranks as scores for the chosen propagation algorithm. Gene expression values from RNA-seq experiments could easily be investigated in the same manner, along with protein abundance data, mutation data or epigenetic data. Finally, there already exist different approaches for applying propagation algorithms to detect network modules. Here we have chosen to use the HotNet2 algorithm, but several others, like the ones in the review by Cowen et al. (2017), might also be considered.

In our work we focus on anthracyclines, a group of commonly used chemotherapy drugs. We used the data that are available in the DrugMatrix (Ganter et al., 2005) database and applied our workflow (Figure 1). This workflow could easily be applied to other data resources as well as other groups of drugs. Some previous works have already been developed to analyze toxicogenomics data from



DrugMatrix and were applied for identifying different types of drug induced toxicities. For example, Tawa et al. (2014) characterized liver induced drug toxicity by identifying gene co-expression modules that are associated with a toxic response. They defined these gene modules using six different methods, including Pearson correlations and PPI information. A similar approach was also applied to identify gene co-expression modules for kidney induced drug toxicity (AbdulHameed et al., 2016). In another work, AbdulHameed et al. (2014) also tried to identify liver induced drug toxicity by integrating toxicogenomics

data with pathway and PPI network information. They performed a differential expression analysis and identified relevant gene modules by applying the KeyPathwayMiner (Alcaraz et al., 2012) algorithm. Other network based approaches have also been suggested for the analysis of toxicogenomics data from the DrugMatrix database. For instance, Sutherland et al. (2017) have constructed gene co-expression networks using WGCNA (Zhang and Horvath, 2005) and associated modules with different drug toxicity phenotypes. Mulas et al. (2017) compiled a pipeline for network comparison and used it to identify drugs with similar

toxicity profiles. In our workflow, we chose to apply a network propagation algorithm that is based on a random walk model. We showed that this approach allows for the identification of drug toxicity modules that are highly enriched in functional information and provide new insights into the toxic causing mechanisms.

Gene expression signatures have been associated with toxicity phenotypes with the concept of phenotypic anchoring (Paules, 2003). Here, the idea is that specific signatures emerge over time and dose that can be related to distinguishable phenotypes. We have observed that, for example the number of DEGs in DOX and IDA at MTDs reflect previously observed differences in the toxicity of both compounds. Additionally, when comparing enrichment scores in heart-related diseases pathways, DOX appears as the most toxic compound followed by EPI, while IDA and DAU show basically no enrichment in these pathways (Supplementary Figure 4).

Associating genotype with phenotype, and specifically predicting a toxic phenotype that rises due to drug treatment, still remains an intricate challenge. Integrating experimental data with prior knowledge in the form of biological networks, as suggested in our work, is a suitable step when trying to describe the molecular effects of drug treatments. However, there is still much to be improved. The PPI networks still hold a high bias in interactions due to annotation (Schramm et al., 2013; Luecken et al., 2018) and will keep getting refined as our understanding of the biological systems increases. Better experimental techniques become more and more available, and data from those will need to be integrated for an even more comprehensive analysis (Hasin et al., 2017; Yan et al., 2017;

Karczewski and Snyder, 2018). And finally, better computational approaches for differentiating between cases and controls, as well as for analyzing big networks such as PPIs, are still to be developed.

AUTHOR CONTRIBUTIONS

GB developed the workflow, performed parts of the data analysis, and wrote the manuscript. RH conceived the study, performed parts of the data analysis, and wrote the manuscript.

FUNDING

The study was in part funded by the European Commission under its 7th Framework Program (Grant HeCaToS, 602156).

ACKNOWLEDGMENTS

We thank Christopher Hardt for assistance in data analysis and Atanas Kamburov for work with ConsensusPathDB.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2018.00484/full#supplementary-material>

REFERENCES

- AbdulHameed, M. D. M., Ippolito, D. L., Stallings, J. D., and Wallqvist, A. (2016). Mining kidney toxicogenomic data by using gene co-expression modules. *BMC Genomics* 17:790. doi: 10.1186/s12864-016-3143-y
- AbdulHameed, M. D. M., Tawa, G. J., Kumar, K., Ippolito, D. L., Lewis, J. A., Stallings, J. D., et al. (2014). Systems level analysis and identification of pathways and networks associated with liver fibrosis. *PLoS One* 9:e112193. doi: 10.1371/journal.pone.0112193
- Alcaraz, N., Friedrich, T., Kötzing, T., Krohmer, A., Müller, J., Pauling, J., et al. (2012). Efficient key pathway mining: combining networks and OMICS data. *Integr. Biol.* 4, 756–764. doi: 10.1039/c2ib00133k
- Andersen, M. E., Clewell, H. J. III, Bermudez, E., Willson, G. A., and Thomas, R. S. (2008). Genomic signatures and dose-dependent transitions in nasal epithelial responses to inhaled formaldehyde in the rat. *Toxicol. Sci.* 105, 368–383. doi: 10.1093/toxsci/kfn097
- Bader, G. D., Cary, M. P., and Sander, C. (2006). Pathguide: a pathway resource list. *Nucleic Acids Res.* 34, D504–D506. doi: 10.1093/nar/gkj126
- Barabasi, A. L., and Oltvai, Z. N. (2004). Network biology: understanding the cell's functional organization. *Nat. Rev. Genet.* 5, 101–113. doi: 10.1038/nrg1272
- Burridge, P. W., Keller, G., Gold, J. D., and Wu, J. C. (2012). Production of de novo cardiomyocytes: human pluripotent stem cell differentiation and direct reprogramming. *Cell Stem Cell* 10, 16–28. doi: 10.1016/j.stem.2011.12.013
- Conesa, A., Nueda, M. J., Ferrer, A., and Talon, M. (2006). maSigPro: a method to identify significantly differential expression profiles in time-course microarray experiments. *Bioinformatics* 22, 1096–1102. doi: 10.1093/bioinformatics/bt1056
- Cowen, L., Ideker, T., Raphael, B. J., and Sharan, R. (2017). Network propagation: a universal amplifier of genetic associations. *Nat. Rev. Genet.* 18, 551–562. doi: 10.1038/nrg.2017.38
- Dai, M., Wang, P., Boyd, A. D., Kostov, G., Athey, B., Jones, E. G., et al. (2005). Evolving gene/transcript definitions significantly alter the interpretation of GeneChip data. *Nucleic Acids Res.* 33:e175. doi: 10.1093/nar/gni179
- Ganter, B., Tugendreich, S., Pearson, C. I., Ayanoglu, E., Baumhueter, S., Bostian, K. A., et al. (2005). Development of a large-scale chemogenomics database to improve drug candidate selection and to understand mechanisms of chemical toxicity and action. *J. Biotechnol.* 119, 219–244. doi: 10.1016/j.jbiotec.2005.03.022
- Geisberg, C. A., and Sawyer, D. B. (2010). Mechanisms of anthracycline cardiotoxicity and strategies to decrease cardiac damage. *Curr. Hypertens. Rep.* 12, 404–410. doi: 10.1007/s11906-010-0146-y
- Gentry, J. (2017). “*gcrma: Background Adjustment Using Sequence Information*.”.
- Goldberg, D. S., and Roth, F. P. (2003). Assessing experimentally derived interactions in a small world. *Proc. Natl. Acad. Sci. U.S.A.* 100, 4372–4376. doi: 10.1073/pnas.0735871100
- Gusenleitner, D., Auerbach, S. S., Melia, T., Gomez, H. F., Sherr, D. H., and Monti, S. (2014). Genomic models of short-term exposure accurately predict long-term chemical carcinogenicity and identify putative mechanisms of action. *PLoS One* 9:e102579. doi: 10.1371/journal.pone.0102579
- Gustafsson, M., Nestor, C. E., Zhang, H., Barabasi, A. L., Baranzini, S., Brunak, S., et al. (2014). Modules, networks and systems medicine for understanding disease and aiding diagnosis. *Genome Med.* 6:82. doi: 10.1186/s13073-014-0082-6
- Hamidi Asl, L., Liepnieks, J. J., Hamidi Asl, K., Uemichi, T., Moulin, G., Desjoyaux, E., et al. (1999). Hereditary amyloid cardiomyopathy caused by a variant apolipoprotein A1. *Am. J. Pathol.* 154, 221–227. doi: 10.1016/S002-9440(10)65268-6
- Hardt, C., Beber, M. E., Rasche, A., Kamburov, A., Hebels, D. G., Kleinjans, J. C., et al. (2016). ToxDB: pathway-level interpretation of drug-treatment data. *Database* 2016:baw052. doi: 10.1093/database/baw052

- Hartung, T. (2009). Toxicology for the twenty-first century. *Nature* 460, 208–212. doi: 10.1038/460208a
- Hasin, Y., Seldin, M., and Lusk, A. (2017). Multi-omics approaches to disease. *Genome Biol.* 18:83. doi: 10.1186/s13059-017-1215-1
- Hendrickx, D. M., Aerts, H. J., Caiment, F., Clark, D., Ebbels, T. M., Evelo, C. T., et al. (2015). diXa: a data infrastructure for chemical safety assessment. *Bioinformatics* 31, 1505–1507. doi: 10.1093/bioinformatics/btu827
- Herwig, R., Hardt, C., Lienhard, M., and Kamburov, A. (2016). Analyzing and interpreting genome data at the network level with ConsensusPathDB. *Nat. Protoc.* 11, 1889–1907. doi: 10.1038/nprot.2016.117
- Hinton, R. B., Adelman-Brown, J., Witt, S., Krishnamurthy, V. K., Osinska, H., Sakthivel, B., et al. (2010). Elastin haploinsufficiency results in progressive aortic valve malformation and latent valve disease in a mouse model. *Circ. Res.* 107, 549–557. doi: 10.1161/CIRCRESAHA.110.221358
- Holmgren, G., Synnergren, J., Bogestal, Y., Ameen, C., Akesson, K., Holmgren, S., et al. (2015). Identification of novel biomarkers for doxorubicin-induced toxicity in human cardiomyocytes derived from pluripotent stem cells. *Toxicology* 328, 102–111. doi: 10.1016/j.tox.2014.12.018
- Kamburov, A., Grossmann, A., Herwig, R., and Stelzl, U. (2012a). Cluster-based assessment of protein-protein interaction confidence. *BMC Bioinformatics* 13:262. doi: 10.1186/1471-2105-13-262
- Kamburov, A., Stelzl, U., and Herwig, R. (2012b). IntScore: a web tool for confidence scoring of biological interactions. *Nucleic Acids Res.* 40, W140–W146. doi: 10.1093/nar/gks492
- Kamburov, A., Stelzl, U., Lehrach, H., and Herwig, R. (2013). The ConsensusPathDB interaction database: 2013 update. *Nucleic Acids Res.* 41, D793–D800. doi: 10.1093/nar/gks1055
- Kamburov, A., Wierling, C., Lehrach, H., and Herwig, R. (2009). ConsensusPathDB—a database for integrating human functional interaction networks. *Nucleic Acids Res.* 37, D623–D628. doi: 10.1093/nar/gkn698
- Kanehisa, M., Goto, S., Sato, Y., Furumichi, M., and Tanabe, M. (2012). KEGG for integration and interpretation of large-scale molecular data sets. *Nucleic Acids Res.* 40, D109–D114. doi: 10.1093/nar/gkr988
- Karczewski, K. J., and Snyder, M. P. (2018). Integrative omics for health and disease. *Nat. Rev. Genet.* 19, 299–310. doi: 10.1038/nrg.2018.4
- Kohonen, P., and Parkkinen, J. A. (2017). A transcriptomics data-driven gene space accurately predicts liver cytopathology and drug-induced liver injury. *Nat. Commun.* 8:15932. doi: 10.1038/ncomms15932
- Kuchaiev, O., Rasajski, M., Higham, D. J., and Przulj, N. (2009). Geometric de-noising of protein-protein interaction networks. *PLoS Comput. Biol.* 5:e1000454. doi: 10.1371/journal.pcbi.1000454
- Leiserson, M. D., Vandin, F., Wu, H. T., Dobson, J. R., Eldridge, J. V., Thomas, J. L., et al. (2015). Pan-cancer network analysis identifies combinations of rare somatic mutations across pathways and protein complexes. *Nat. Genet.* 47, 106–114. doi: 10.1038/ng.3168
- Lenneman, C. G., and Sawyer, D. B. (2016). Cardio-oncology: an update on cardiotoxicity of cancer-related treatment. *Circ. Res.* 118, 1008–1020. doi: 10.1161/CIRCRESAHA.115.303633
- Liebler, D. C., and Guengerich, F. P. (2005). Elucidating mechanisms of drug-induced toxicity. *Nat. Rev. Drug Discov.* 4, 410–420. doi: 10.1038/nrd1720
- Luecken, M. D., Page, M. J. T., Crosby, A. J., Mason, S., Reinert, G., and Deane, C. M. (2018). CommWalker: correctly evaluating modules in molecular networks in light of annotation bias. *Bioinformatics* 34, 994–1000. doi: 10.1093/bioinformatics/btx706
- Luo, W., Pant, G., Bhavnasi, Y. K., Blanchard, S. G. Jr., and Brouwer, C. (2017). Pathview Web: user friendly pathway visualization and data integration. *Nucleic Acids Res.* 45, W501–W508. doi: 10.1093/nar/gkx372
- Maillet, A., Tan, K., Chai, X., Sadananda, S. N., Mehta, A., Ooi, J., et al. (2016). Modeling doxorubicin-induced cardiotoxicity in human pluripotent stem cell derived-cardiomyocytes. *Sci. Rep.* 6:25333. doi: 10.1038/srep25333
- Marco-Ramell, A., Palau-Rodriguez, M., Alay, A., Tulipani, S., Urpi-Sarda, M., Sanchez-Pla, A., et al. (2018). Evaluation and comparison of bioinformatic tools for the enrichment analysis of metabolomics data. *BMC Bioinformatics* 19:1. doi: 10.1186/s12859-017-2006-0
- Martha, V. S., Liu, Z., Guo, L., Su, Z., Ye, Y., Fang, H., et al. (2011). Constructing a robust protein-protein interaction network by integrating multiple public databases. *BMC Bioinformatics* 12(Suppl. 10):S7. doi: 10.1186/1471-2105-12-S10-S7
- Matthews, L., Gopinath, G., Gillespie, M., Caudy, M., Croft, D., De Bono, B., et al. (2009). Reactome knowledgebase of human biological pathways and processes. *Nucleic Acids Res.* 37, D619–D622. doi: 10.1093/nar/gkn863
- McGillivray, P., Clarke, D., Meyerson, W., Zhang, J., Lee, D., Gu, M., et al. (2018). Network analysis as a grand unifier in biomedical data science. *Annu. Rev. Biomed. Data Sci.* 1, 153–180. doi: 10.1038/ncomms10031
- McGowan, J. V., Chung, R., Maulik, A., Piotrowska, I., Walker, J. M., and Yellon, D. M. (2017). Anthracycline chemotherapy and cardiotoxicity. *Cardiovasc. Drugs Ther.* 31, 63–75. doi: 10.1007/s10557-016-6711-0
- Mei, N., Fuscoe, J. C., Lobenhofer, E. K., and Guo, L. (2010). Application of microarray-based analysis of gene expression in the field of toxicogenomics. *Methods Mol. Biol.* 597, 227–241. doi: 10.1007/978-1-60327-389-3_16
- Mills, P. J., Ancoli-Israel, S., Parker, B., Natarajan, L., Hong, S., Jain, S., et al. (2008). Predictors of inflammation in response to anthracycline-based chemotherapy for breast cancer. *Brain Behav. Immun.* 22, 98–104. doi: 10.1016/j.bbi.2007.07.001
- Mulas, F., Li, A., Sherr, D. H., and Monti, S. (2017). Network-based analysis of transcriptional profiles from chemical perturbations experiments. *BMC Bioinformatics* 18:130. doi: 10.1186/s12859-017-1536-9
- Nystrom-Persson, J., Natsume-Kitatani, Y., Igarashi, Y., Satoh, D., and Mizuguchi, K. (2017). Interactive Toxicogenomics: gene set discovery, clustering and analysis in Toxygates. *Sci. Rep.* 7:1390. doi: 10.1038/s41598-017-01500-1
- Paules, R. (2003). Phenotypic anchoring: linking cause and effect. *Environ. Health Perspect.* 111, A338–A339. doi: 10.1289/ehp.111-a338
- Platel, D., Pouna, P., Bonoron-Adele, S., and Robert, J. (1999). Comparative cardiotoxicity of idarubicin and doxorubicin using the isolated perfused rat heart model. *Anticancer Drugs* 10, 671–676. doi: 10.1097/00001813-199908000-00007
- Raschi, E., Vasina, V., Ursino, M. G., Boriani, G., Martoni, A., and De Ponti, F. (2010). Anticancer drugs and cardiotoxicity: insights and perspectives in the era of targeted therapy. *Pharmacol. Ther.* 125, 196–218. doi: 10.1016/j.pharmthera.2009.10.002
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). limma powers differential expression analyses for RNA-seq and microarray studies. *Nucleic Acids Res.* 43:e47. doi: 10.1093/nar/gkv007
- Rueda-Zarate, H. A., Imaz-Rosshandler, I., Cardenas-Ovando, R. A., Castillo-Fernandez, J. E., Noguez-Monroy, J., and Rangel-Escareno, C. (2017). A computational toxicogenomics approach identifies a list of highly hepatotoxic compounds from a large microarray database. *PLoS One* 12:e0176284. doi: 10.1371/journal.pone.0176284
- Sauter, K. A., Wood, L. J., Wong, J., Iordanov, M., and Magun, B. E. (2011). Doxorubicin and daunorubicin induce processing and release of interleukin-1beta through activation of the NLRP3 inflammasome. *Cancer Biol. Ther.* 11, 1008–1016. doi: 10.4161/cbt.11.12.15540
- Schaefer, C. F., Anthony, K., Krupa, S., Buchoff, J., Day, M., Hannay, T., et al. (2009). PID: the Pathway Interaction Database. *Nucleic Acids Res.* 37, D674–D679. doi: 10.1093/nar/gkn653
- Schramm, S. J., Jayaswal, V., Goel, A., Li, S. S., Yang, Y. H., Mann, G. J., et al. (2013). Molecular interaction networks for the analysis of human disease: utility, limitations, and considerations. *Proteomics* 13, 3393–3405. doi: 10.1002/pmic.201200570
- Steinherz, L. J., Steinherz, P. G., Tan, C. T., Heller, G., and Murphy, M. L. (1991). Cardiac toxicity 4 to 20 years after completing anthracycline therapy. *JAMA* 266, 1672–1677. doi: 10.1001/jama.1991.03470120074036
- Stillitano, F., Karakikes, I., Costa, K. D., Fish, K., Hajjar, R. J., and Hulot, J.-S. (2012). Preclinical animal models for testing iPSC/ESC-based heart therapy. *Drug Discov. Today* 9, e229–e236.
- Sutherland, J. J., Webster, Y. W., Willy, J. A., Searfoss, G. H., Goldstein, K. M., Irizarry, A. R., et al. (2017). Toxicogenomic module associations with pathogenesis: a network-based approach to understanding drug toxicity. *Pharmacogenomics J.* 18:377. doi: 10.1038/tpj.2017.17
- Tawa, G. J., Abdulhameed, M. D. M., Yu, X., Kumar, K., Ippolito, D. L., Lewis, J. A., et al. (2014). Characterization of chemically induced liver injuries using gene co-expression modules. *PLoS One* 9:e107230. doi: 10.1371/journal.pone.0107230

- Truong, J., Yan, A. T., Cramarossa, G., and Chan, K. K. (2014). Chemotherapy-induced cardiotoxicity: detection, prevention, and management. *Can. J. Cardiol.* 30, 869–878. doi: 10.1016/j.cjca.2014.04.029
- Uehara, T., Ono, A., Maruyama, T., Kato, I., Yamada, H., Ohno, Y., et al. (2010). The Japanese toxicogenomics project: application of toxicogenomics. *Mol. Nutr. Food Res.* 54, 218–227. doi: 10.1002/mnfr.200900169
- Vandin, F., Clay, P., Upfal, E., and Raphael, B. J. (2012). Discovery of mutated subnetworks associated with clinical data in cancer. *Pac. Symp. Biocomput.* 17, 55–66. doi: 10.1142/9789814366496_0006
- Vandin, F., Upfal, E., and Raphael, B. J. (2011). Algorithms for detecting significantly mutated pathways in cancer. *J. Comput. Biol.* 18, 507–522. doi: 10.1089/cmb.2010.0265
- Vidal, M., Cusick, M. E., and Barabasi, A. L. (2011). Interactome networks and human disease. *Cell* 144, 986–998. doi: 10.1016/j.cell.2011.02.016
- Walhout, A. J., and Vidal, M. (2001). Protein interaction maps for model organisms. *Nat. Rev. Mol. Cell Biol.* 2, 55–62. doi: 10.1038/35048107
- Wang, K. C., Botting, K. J., Padhee, M., Zhang, S., Mcmillen, I. C., Suter, C. M., et al. (2012). Early origins of heart disease: low birth weight and the role of the insulin-like growth factor system in cardiac hypertrophy. *Clin. Exp. Pharmacol. Physiol.* 39, 958–964. doi: 10.1111/j.1440-1681.2012.05743.x
- Xia, K., Dong, D., and Han, J. D. (2006). IntNetDB v1.0: an integrated protein-protein interaction network database generated by a probabilistic model. *BMC Bioinformatics* 7:508.
- Yan, J., Risacher, S. L., Shen, L., and Saykin, A. J. (2017). Network approaches to systems biology analysis of complex disease: integrative methods for multi-omics data. *Brief Bioinform.* [Epub ahead of print]. doi: 10.1093/bib/bbx066
- Yates, A., and Akanni, W. (2016). Ensembl 2016. *Nucleic Acid Res.* 44, D710–D716. doi: 10.1093/nar/gkv1157
- Yildirimman, R., Brolen, G., Vilardell, M., Eriksson, G., Synnergren, J., Gmuender, H., et al. (2011). Human embryonic stem cell derived hepatocyte-like cells as a tool for in vitro hazard assessment of chemical carcinogenicity. *Toxicol. Sci.* 124, 278–290. doi: 10.1093/toxsci/kfr225
- Yu, G., Li, F., Qin, Y., Bo, X., Wu, Y., and Wang, S. (2010). GOSemSim: an R package for measuring semantic similarity among GO terms and gene products. *Bioinformatics* 26, 976–978. doi: 10.1093/bioinformatics/btq064
- Zhang, B., and Horvath, S. (2005). A general framework for weighted gene co-expression network analysis. *Stat. Appl. Genet. Mol. Biol.* 4:Article17. doi: 10.2202/1544-6115.1128
- Zheng, G., and Huang, T. (2018). The Reconstruction and analysis of gene regulatory networks. *Methods Mol. Biol.* 1754, 137–154. doi: 10.1007/978-1-4939-7717-8_8

Conflict of Interest Statement: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2018 Barel and Herwig. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.