ABSTRACT: Transcriptomic approaches can give insight into molecular mechanisms underlying chemical toxicity and are increasingly being used as part of toxicological assessments. To aid the interpretation of transcriptomic data, we have developed a systems toxicology method that relies on a computable biological network model. We created the first network model describing cardiotoxicity in zebrafish larvae—a valuable emerging model species in testing cardiotoxicity associated with drugs and chemicals. The network is based on scientific literature and represents hierarchical molecular pathways that lead from receptor activation to cardiac pathologies. To test the ability of our approach to detect cardiotoxic outcomes from transcriptomic data, we have selected three publicly available data sets that reported chemically induced heart pathologies in zebrafish larvae for five different chemicals. Network-based analysis detected cardiac perturbations for four out of five chemicals tested, for two of them using transcriptomic data collected up to 3 days before the onset of a visible phenotype. Additionally, we identified distinct molecular pathways that were activated by the different chemicals. The results demonstrate that the proposed integrational method can be used for evaluating the effects of chemicals on the zebrafish cardiac function and, together with observed cardiac apical end points, can provide a comprehensive method for connecting molecular events to organ toxicity. The computable network model is freely available and may be used to generate mechanistic hypotheses and quantifiable perturbation values from any zebrafish transcriptomic data.

INTRODUCTION

Cardiotoxicity is an adverse effect associated with many drugs and environmental pollutants. This raises the need for good experimental models for assessing the potential cardiotoxicity associated with chemicals such as drugs and pesticides. Cell-based in vitro assays are fast and inexpensive. However, they have limited predictive power because of their inherent biological simplicity. For example, drug effects in vivo are governed not only by molecular activity but also by absorption, distribution, metabolism, and excretion of the drug, which are difficult to study in cultured cells. Although rodent animal models are extremely valuable tools in toxicology, chemical assessment in mammals is time-consuming and relatively expensive.

The zebrafish (Danio rerio) is an alternative animal model that can help address the above limitations. The zebrafish is a small subterminal fish that has several characteristics useful for toxicological screening: It is small and simple to maintain and amenable to large-scale studies owing to its high fecundity, transparent larval stages, and well-defined and rapid external development. Additionally, zebrafish larvae are not subject to animal regulations for up to 5 days postfertilization (dpf).

These advantages have established the zebrafish as a prominent model organism in environmental toxicology. Remarkably, zebrafish and humans exhibit similar responses to cardiotoxic drugs, signifying that the zebrafish is a useful model for studying human chemical cardiotoxicity. Phenotypic assessment of cardiotoxicity in zebrafish is relatively straightforward. Its heart morphology is readily observable under a dissecting microscope, and commercially available software can be used to measure the heart rate. Human and zebrafish heart rates are more similar to each other than that of mouse (human, 60–100 beats per minute (bpm); mouse, 300–600 bpm; zebrafish, 120–180 bpm). Reflecting this, zebrafish and humans show similar electrocardiogram (ECG) recordings. Finally, passive diffusion of oxygen sustains the larvae for up to 4 to 5 dpf, such that blood circulation up to this age is not required.
for survival. This allows assessment of cardiac toxicity independently from oxygen deprivation, something not possible in mammals. Because of this, the zebrafish is not only valuable in environmental toxicology but also a faithful model for studying chemical and drug toxicity in humans.

Classical toxicology methods have traditionally been used for assessing chemical toxicity. These methods are largely descriptive, because they report dose-dependent adverse outcomes. However, when studying chemical toxicity, understanding the molecular events that lead to an adverse outcome is as important as understanding the outcome itself. Mechanistic knowledge of toxicity allows in silico modeling, extrapolation to chemicals with similar modes of action, prediction of chronic exposure effects, exposure in other species (including humans), development of therapeutic strategies, and, ultimately, regulatory decision-making. Mechanistic insights into toxicity are particularly pertinent when using the zebrafish as a model. In toxicology and ecotoxicology, this knowledge would better inform relevance to toxicity for other fish species as well as humans.

Transcriptomic approaches have become a valuable tool for generating hypotheses on the molecular mechanisms of action underlying adverse chemical effects. However, interpretation of single-gene-expression values can be challenging in a toxicological context, because abundance of mRNA is not directly linked to adverse outcomes. Additionally, molecular events leading up to toxicological outcomes are not always apparent from static gene expression data, necessitating time-resolved experiments. This makes identification of mode of action of chemicals more time-consuming and costly.

We have developed a systems toxicology method with the aim of advancing the zebrafish as a model in (eco)toxicology specifically and improving the interpretation of transcriptomic data in a toxicological context in general. Our method considers the impact of a toxicant on the zebrafish heart as a whole system instead of relying on interpretation of single-gene-expression values. The approach relies on a causal biological network model—a collection of computable statements curated from scientific literature describing connected molecular pathways. Nodes within the network represent entities, at diverse levels of biological organization, for example, mRNAs, proteins, protein activities, biological processes, and pathologies. The nodes are connected by directed and signed relationships (i.e., upregulation or inhibition), giving the network a cause-and-effect topology. Pre-existing gene expression signatures associated with specific nodes in the network are then compared with any given compound-induced gene-expression values to determine the activity of the node. This approach, therefore, treats changes in the abundance of transcripts as a downstream response to node activity, without assuming that these transcripts produce active proteins. In this way, a single transcriptomic measurement (instead of many measurements at different biological levels) can be used to determine the activity of the whole network. Furthermore, because the network is directed, it is possible to follow the biological response induced by a compound from molecular initiating events to adverse outcomes, given that both are part of the network.

While causal biological networks have been previously built for human, rat, and mouse molecular processes, the cardiotoxicity network introduced in this study is the first such network for zebrafish. To develop the network, we built a zebrafish namespace (based on the Zebrafish Information Network [ZFIN]) as a reference for conventional naming of genes, transcripts, and proteins. We then reviewed the cardiotoxicity literature and used the Biological Expression Language (BEL) to construct a network describing cardiotoxic events in zebrafish. In this report, we describe the individual steps required to build the network and demonstrate how the network can be used for interpreting transcriptomic data from chemically treated samples to predict adverse cardiac outcomes.

## EXPERIMENTAL PROCEDURES

### Causal Biological Networks

Causal biological network models developed in our approach are constructed from qualitative causal relationships curated from relevant literature by using BEL version 1.0 (https://github.com/OpenBEL/language) (Figure S1). BEL is a tool for representing published biological findings in a computable form by coding molecular relationships as BEL statements. Each BEL statement is a triple of a subject and an object (termed nodes) and a relationship between them (called an edge). Nodes in a BEL statement represent biological entities such as RNA, protein, protein activity, biological process, or pathology. Edges represent positive or negative regulation. Additionally, each BEL statement is annotated with metadata to facilitate future verification: the publication used to generate the statement, species, tissue, and experimental methods. BEL statements were generated with the assistance of the BEL Information Extraction workFlow (BELIEF) semiautomated curation platform. To detect zebrafish gene and protein names in scientific text and encode them in BEL, we integrated two new components in BELIEF, a named entity recognition module and a zebrafish gene namespace for BEL. For the named entity recognition module, we used the software ProMiner, which facilitates detection of zebrafish gene and protein names in text through a dictionary- and rule-based approach. To build the dictionary for ProMiner and the BEL namespace, we gathered zebrafish gene names and IDs from ZFIN, which serves as a zebrafish model organism database. It offers the Genetic Markers database, which includes identifiers, symbols, and names for various genetic objects such as genes, pseudogenes, and transcripts. From this database, we extracted all genes and pseudogenes, along with their unique identifiers. We also employed the mappings between zebrafish genetic markers and the UniProt knowledgebase that are provided by the ZFIN Marker and UniProtKB database.

BEL statements were compiled into a cohesive network by using the open-source compilation framework OpenBEL 3.0.0 (https://github.com/OpenBEL/openbel-framework). The resulting network was visualized by using the Cytoscape software—an open-source platform for representation and analysis of complex networks. An interactive version of the network model is accessible on the Causal Biological Network Database (http://causalbionet.com/Home/NetworkVisualization/#/1/networkGraph?ID=5ea0344201dd7708605b4e09).

### Network Scoring

To avoid the assumption that mRNA abundance corresponds with protein abundance and activity, we leveraged the “reverse causal reasoning” paradigm. This paradigm relies on prior knowledge of transcripts known to be differentially regulated by a given node. For example, a transcriptomic experiment might indicate that overexpression of a transcription factor leads to upregulation of transcript set X and downregulation of transcript set Y (Figure 1a,b). This gene expression signature may now be used to infer the activity of this transcription factor in any given transcriptomic data set with differential gene expression values. Such nodes, which harbor information about regulation of gene expression, are called inferable nodes (iNodes). Mapping these molecular signatures downstream of the nodes in the network generates the downstream transcript layer. Transcripts that constitute the transcript layer are directionally connected to their corresponding node in the functional layer but not to each other (Figure 1b). iNodes are built largely by using data collected from publicly available transcriptomic data sets.

Inferred differential activity of individual iNodes (determined by similarity between gene expression measured after chemical exposure and gene expression encoded in the transcript layer for each node) in the causal network allows quantification of network perturbation

https://doi.org/10.1021/acs.chemrestox.0c00095

2551
Contradictions between node activities and edge signs produce a high network perturbation amplitude (NPA) score. High consistency between signed edges and inferred activities of the blue nodes are inferred to be downregulated; gray nodes are unchanged. Red nodes are inferred to be upregulated; values, node C would be inferred to be activated because of treatment 1, there is a close agreement between transcript abundances associated with initiating events (such as receptor activation) at the top and consistent with the neighboring nodes in the functional layer. This analysis has been shown to be a useful way for interpreting the biological basis of a given perturbation. In addition to identification of the important regions of the network responsible for the observed impact, LN analysis indicates the directionality of the impact on each node.

### RESULTS

#### Zebrafish Cardiotoxicity Network

The zebrafish cardiotoxicity network is a compilation of computable statements curated from peer-reviewed articles and encoded in BEL. We selected articles that described cardiotoxic outcomes in zebrafish larvae brought about by chemical exposure and gain- or loss-of-function analysis. For this purpose, we did not take into account the impact factor of the journal or any other measure for ranking the articles, but relied on careful manual evaluation of the articles. Of the nine adverse cardiac outcomes initially selected, the following six were removed from the final analysis because of sparse mechanistic data: endomyocardial fibrosis, cardiac arrhythmia, Long QT syndrome, heart failure, hyperplasia, and heart arrest. The remaining cardiotoxic outcomes, with sufficient data available in the literature, were hypertrophy, edema, and bradycardia. Cardiac hypertrophy is a pathological enlargement of the heart muscle. Pericardial edema is caused by accumulation of fluid around the heart resulting in enlargement of the pericardial sac. Bradycardia is characterized by an abnormally slow heart rate. After establishing molecular events that lead to cardiac toxicity (e.g., “activation of sh2r increases cardiac edema”), literature describing upstream and downstream events was identified and curated (e.g., “activation of sh2r increases ptgs2b and “ptgs2b increases cardiac edema”). Evidence was also included from zebrafish studies that relied on data previously shown in other model organisms (e.g., rat or mouse). These data were largely used to fill gaps between nodes corresponding to evolutionarily conserved pathways, such as the mitogen-activated protein kinase (MAPK) cascade. This approach resulted in the curation of 81 articles (listed in Table S1) to create the causal cardiotoxicity network.

The network contains 232 nodes connected by 348 edges (Figure 2a). Apart from the three pathologies already mentioned, the nodes represent 99 proteins, 75 protein activities, 15 mRNAs, 15 protein complexes, 9 metabolites, 8 biological processes, 5 protein families, and 3 microRNAs. Each edge has associated metadata identifying the evidence, species, Medical Subject Headings (MESH) terms, and publication used to create the edge. The network is represented hierarchically, with initiating events (such as receptor activation) at the top and adverse outcomes (such as pathologies) at the bottom. The nodes with the most inward connections are “hypertrophy”, “cardiac edema”, and “cardiac muscle contraction” (Table 1). The nodes with the most outgoing connections are calcium/calmodulin-dependent protein kinase II beta 2 (cank2b2), histone deacetylase 9b (hdac9b), and calcium ion (Table 1).

Many molecular pathways known to play a role in cardiac toxicity are represented in the network (Figure 2b). For example, the MAPK cascade is a highly conserved pathway that regulates fundamental cellular processes such as proliferation.
and differentiation. The signal within this pathway propagates through sequential phosphorylation and activation of kinases Raf, Mek1/2, and Erk1/2, \(^{25}\) improper activation of this cascade is implicated in cardiac hypertrophy. \(^{26}\) mTOR (mechanistic target of rapamycin) acts downstream of PI3K (phosphatidylinositol 3-kinase) and PKB (protein kinase B), also known as

![Figure 2](image)

**Figure 2.** Functional layer of the zebrafish cardiotoxicity network. (a) Causal biological network representing molecular events that lead to cardiotoxic outcomes in zebrafish. The network is represented hierarchically, with initiating events at the top and pathological end points at the bottom. (b) Selected pathways extracted from the full network. Dotted line represents an indirect connection in the full network.

**Table 1.** Nodes with the Most Outward (Edges Out) and Inward (Edges In) Connections

<table>
<thead>
<tr>
<th>node name</th>
<th>edges out</th>
<th>node name</th>
<th>edges in</th>
</tr>
</thead>
<tbody>
<tr>
<td>act(p(ZFIN:camk2b2))</td>
<td>7</td>
<td>path(MESH:Cardiac hypertrophy)</td>
<td>17</td>
</tr>
<tr>
<td>act(p(ZFIN:hdac9b))</td>
<td>6</td>
<td>path(MESH:Cardiac hypertrophy)</td>
<td>11</td>
</tr>
<tr>
<td>act(p(ZFIN:ptpn11a))</td>
<td>5</td>
<td>bp(GOBP:cardiac muscle contraction)</td>
<td>8</td>
</tr>
<tr>
<td>act(p(ZFIN:ilk))</td>
<td>5</td>
<td>p(ZFIN:akt1,pmod(P))</td>
<td>7</td>
</tr>
<tr>
<td>act(p(ZFIN:akt1))</td>
<td>5</td>
<td>p(SFAM:MAPK Erk1/2 Family,pmod(P))</td>
<td>7</td>
</tr>
</tbody>
</table>
Akt, and regulates cell growth. As such, mTOR signaling is indispensable for normal embryonic cardiac development; however, excessive activation of this pathway is associated with cardiac disease. Calcium plays an essential role in cardiac contraction. Calcium binds its target calmodulin (calm1a) and triggers activation of camk2b (Ca2+/calmodulin-dependent protein kinase II), which, in turn, facilitates the opening of the L-type calcium channel (caecn2). The resultant increase in intracellular calcium levels is necessary for cardiac muscle contraction, as experimental inhibition of components of this pathway causes bradycardia, among other phenotypes. Additionally, the network describes two adverse outcome pathways (AOPs) currently under development in the AOP database AOPwiki (www.aopwiki.org) (Figure 2b): “Aryl hydrocarbon receptor activation leading to early life stage mortality, via reduced VEGF” (AOP 150) and “Aryl hydrocarbon receptor activation leading to early life stage mortality, via increased COX-2” (AOP 21) 35,32

Network Scoring with Public Data Sets. To test the ability of the cardiotoxicity network to reflect *in vivo* outcomes, we sought publicly available transcriptomic data sets with associated descriptive cardiotoxic outcomes in zebrafish larvae. To perform this analysis, we selected data sets GSE55618, GSE44130, GSE61155, and GSE72918 from the Gene Expression Omnibus repository. Data set GSE55618 was included to provide a DMSO negative control, to which the following results were normalized to (Figure 3). All the data sets used for network scoring were independent from the studies used to construct the network.

In the GSE44130 study, the authors evaluated transcriptional responses induced by three polycyclic aromatic hydrocarbons (PAHs), benz[a]anthracene (BAA), dibenzo thiophene (DBT), and pyrene (PYR) in zebrafish embryos at 24 and 48 h postexposure. PAHs are byproducts of combustion present in the environment and are known to cause developmental abnormalities in zebrafish. All three chemicals induced pericardial edema at 5 dpf in the study and can, therefore, be classified as cardiotoxins. Our NPA analysis revealed that treatment with BAA and DBT did not significantly perturb the cardiotoxicity network at 24hpf; however, perturbation of the network was significant at 48hpf (Figure 3). PYR treatment did not result in significant network perturbation at either time point.

In the GSE61155 study, zebrafish larvae were treated with trichloroethylene (TCE), an industrially used chlorinated hydrocarbon of environmental concern. The authors reported a significant decrease in mitochondrial function in the heart of zebrafish larvae treated with TCE. In agreement, the transcriptome of TCE-treated zebrafish showed significant perturbation of the cardiotoxicity network (Figure 3).

In the GSE72918 study, the authors treated zebrafish larvae with sorafenib, an anticancer drug associated with adverse cardiac effects in humans. Zebrafish treated with sorafenib exhibited pericardial edema at 4 dpf. In agreement with the observed phenotype, the cardiotoxicity network was significantly perturbed when scored with this data set (Figure 3).

**Leading Node Analysis.** We performed LN analysis to gain mechanistic insights into the cardiotoxic outcomes induced by the chemical treatments discussed above. As demonstrated in the following section, this approach allows us to visualize signal propagation through the network and define the most impacted areas of the network. LN analysis was performed on significantly perturbed networks corresponding to treatments with BAA for 48 h, DBT for 48 h, sorafenib, and TCE (Table 2).

LN analysis of 48h BAA exposure revealed 33 LNs (Figure 4a). The affected nodes in Figure 4a are colored and highlight the affected areas of the network. We extracted these LNs together with adverse outcome nodes to create a subnetwork that represents cardiotoxicity pathways induced by BAA (Figure 4b). This approach helped identify upregulation of the Akt1/mTOR pathway and downregulation of the MAPK cascade after BAA treatment. Additionally, transcription factors Jun and Stat3 were upregulated, and Creb1a (cAMP responsive element binding protein 1a) was inferred to be downregulated in the BAA-treated group. Nppa (natriuretic peptide A) and Nppb were both inferred to be upregulated by BAA. Bmp2b (bone morphogenetic protein 2b), Bmp4, Smad1 (SMAD family member 1), and Smad5 were both inferred to be upregulated by BAA treatment. Thirty-seven LNs were identified in the 48h DBT-treated sample (Figure 5a). The LN subnetwork highlighted upregulation of nitric oxide (NO), members of the MAPK cascade, Akt1, Bmp/Smad, and Creb1a (Figure 5b).

Twenty-six LNs were identified in the sorafenib-treated group (Figure 6a). The LN subnetwork identified upregulation of NO, calcium, Tomm70a (translocase of outer mitochondrial membrane 70 homologue A), Opal (OPAL mitochondrial dynamin like GTPase), Raf1a (Raf-1 proto-oncogene, serine/threonine kinase a), Creb1a, and Bmp/Smad (Figure 6b).

TCE treatment resulted in 41 LNs (Figure 7a). The LN subnetwork revealed elevated activity of MAPK cascade, reactive oxygen species (ROS), and Creb1a and reduced activity of NO, PI3K/Akt/mTOR signaling, Jun, NADH dehydrogenase complex I, Opal1, Hmox1 (heme oxygenase 1), Vegfa (vascular endothelial growth factor Aa), and Kdr (kinase insert domain receptor like) (Figure 7b).

![Figure 3](https://dx.doi.org/10.1021/acs.chemrestox.0c00095)

**Figure 3.** NPA scores for the indicated chemical treatments. Values are normalized to DMSO treatment (dotted line). Gray bars indicate nonsignificant perturbation. Orange bars and red asterisks indicate significant perturbation. Error bars represent 95% confidence intervals.

[^8]: *p < 0.05,* “O” statistic indicates that permutation of the transcript layer results in a significantly different NPA. “K” statistic indicates that permutation of the functional layer results in a significantly different NPA. NPA — network perturbation amplitude.
The zebra when evaluating adverse effects on humans, and vice versa. 37

The zebrafish is a small vertebrate amenable to high-throughput methods and systems approaches. These characteristics make the zebrafish well suited for toxicological assessment of drugs and chemical pollutants. Understanding the molecular mechanisms behind chemical toxicity might help extrapolate the data to other species, including humans, as well as identify effective preventative or therapeutic measures. To address this need, we developed the first computable causal biological network for describing cardiotoxicity in fish.

**Cardiotoxicity Network and AOPs.** The causal cardiotoxicity network describes our current mechanistic knowledge on three cardiac adverse outcomes in zebrafish. Enlargement of the heart and pericardial area and changes in heartbeat are sensitive indicators of cardiac toxicity in zebrafish larvae. Therefore, these adverse outcomes are often reported as a part of toxicological assessment. Reflecting this, the nodes with the most inward connections are hypertrophy, pericardial edema, and cardiac muscle contraction. The functional layer of the network is quite large (232 nodes and 348 edges) and includes many molecular pathways known to play a role in cardiotoxicity (Figure 2b). These and other pathways share connections are hypertrophy, pericardial edema, and cardiac muscle contraction. The functional layer of the network is quite large (232 nodes and 348 edges) and includes many molecular pathways known to play a role in cardiotoxicity (Figure 2b). These and other pathways share

**DISCUSSION**

In the past decade, it has been increasingly recognized that knowledge and information about chemical toxicity to species not traditionally used in chemical risk assessment can be useful when evaluating adverse effects on humans, and vice versa. 37

The zebrafish is a small vertebrate amenable to high-throughput methods and systems approaches. These characteristics make the zebrafish well suited for toxicological assessment of drugs and chemical pollutants. Understanding the molecular mechanisms behind chemical toxicity might help extrapolate the data to other species, including humans, as well as identify effective preventative or therapeutic measures. To address this need, we developed the first computable causal biological network for describing cardiotoxicity in fish.

**Cardiotoxicity Network and AOPs.** The causal cardiotoxicity network describes our current mechanistic knowledge on three cardiac adverse outcomes in zebrafish. Enlargement of the heart and pericardial area and changes in heartbeat are sensitive indicators of cardiac toxicity in zebrafish larvae. Therefore, these adverse outcomes are often reported as a part of toxicological assessment. Reflecting this, the nodes with the most inward connections are hypertrophy, pericardial edema, and cardiac muscle contraction. The functional layer of the network is quite large (232 nodes and 348 edges) and includes many molecular pathways known to play a role in fish and mammalian cardiotoxicity (Figure 2b). These and other pathways share
Figure 4. BAA 48h LN analysis. (a) Cardiotoxicity network scored with the BAA 48h transcriptome. LNs and their percentage contribution to the NPA are indicated in blue (downregulation) and red (activation). (b) LNs extracted into a subnetwork. Dotted lines connect nodes that are indirectly connected in the main model. BAA—benz[a]anthracene, LNs—leading nodes, NPA—network perturbation amplitude.

day nodes in the network. Such connectedness, unique to large biological networks, might, in the future, facilitate discovery of cross-talk between signaling pathways involved in cardiac toxicity, detection of faithful biomarkers of chemical toxicity, and identification of candidate cardiac drug targets.
The in-depth curation based on literature, directionality of edges between nodes, and inclusion of nodes that represent diverse levels of biological organization make our method similar to the approach taken in developing AOPs.\textsuperscript{38,39}

Figure 5. DBT 48h LN analysis. (a) Cardiotoxicity network scored with the DBT 48h transcriptome. LNs and their percentage contribution to the NPA are indicated in blue (downregulation) and red (activation). (b) LNs extracted into a subnetwork. Dotted lines connect nodes that are indirectly connected in the main model. DBT = dibenzothiophene, NPA = network perturbation amplitude.
Therefore, it is not surprising that some of the AOPs recently released by the OECD (Organisation for Economic Co-operation and Development)—AOP 21 and AOP 150—can also be found in our network.31,32 This similarity makes the curated functional layer of the network a useful resource for future AOP development. In effect, the network may be thought of as a network of putative AOPs that share key events. Pertinently, the concept of AOP networks is an emerging area of AOP research, designed to address the inherent simplicity of individual AOPs,40,41 and our approach is similar to the recently developed quantitative AOP methodologies that can be used to infer chemical toxicity.42,43 Lastly, similarly to AOPs, the network is chemical agnostic. To leverage the network as an AOP resource, we are currently working on automatically extracting useful pathways from receptors or molecular initiating events to adverse outcomes. This is possible by identifying simple paths between relevant nodes, but the challenge is identifying a smaller number of the most toxicologically relevant paths among the myriad of possible paths that the network offers. One way to narrow down these possibilities in the future.

Figure 6. Sorafenib LN analysis. (a) Cardiotoxicity network scored with the sorafenib transcriptome. LNs and their percentage contribution to the NPA are indicated in blue (downregulation) and red (activation). (b) LNs extracted into a subnetwork. Dotted lines connect nodes that are indirectly connected in the main model. LNs—leading nodes, NPA—network perturbation amplitude.
is to identify most frequently detected LNs that play a role in chemical cardiotoxicity and use them as a starting point for AOP generation.

Network Scoring. To test the utility of our approach in predicting cardiotoxicity, we selected several studies in which researchers evaluated cardiac phenotypes and measured gene expression in zebrafish for known cardiotoxicants. We sought to test whether, and under what conditions, our approach is able to infer cardiotoxicity, and which molecular pathways contributed the most to perturbation of the causal network.

Network Scoring with PAHs. Goodale et al. exposed zebrafish to BAA, DBT, or PYR and reported pericardial edema.
at 120 hpf. The authors measured gene expression at 24 and 48 hpf in order to evaluate gene expression as an early marker of cardiotoxicity. Our approach was not successful in identifying cardiotoxicity after PYR treatment at either time point. The lower body burden in PYR treatment compared to the other PAHs in the data set is a likely factor in this. We could infer cardiotoxicity after BAA and DBT treatment at 48 h but not at 24 h. The zebrafish heart just begins to function at 24 h, and the transcriptome is highly dynamic during heart morphogenesis. It is probable that, at these early stages, changes in gene expression are largely driven by heart morphogenesis, obscuring any cardiotoxic effects. However, at 48 h–3 days before the cardiac phenotype was reported—our systems toxicology approach predicted cardiotoxic outcomes in response to BAA and DBT treatment, indicating that the method is more effective in predicting effects for longer exposure and greater maturity.

We performed LN analysis to gain insights into the underlying mechanisms of cardiotoxicity induced by the above chemicals. Goodale et al. noted that, although the cardiac phenotype was present in both BAA and DBT treatments, the transcriptional responses to BAA and DBT treatment were largely distinct despite some similarities. Reflecting this observation, our LN analysis identified nodes and pathways both similarly and differentially regulated in response to the two treatments. For example, Akt activity was upregulated in response to both BAA and DBT treatments. Akt/mTOR signaling promotes protein synthesis and cell growth and proliferation under normal conditions, and the lack of mTOR in embryonic murine heart synthesis and cell growth and proliferation under normal conditions, example, Akt activity was upregulated in response to both BAA and DBT treatment. Akt/mTOR signaling promotes protein synthesis and cell growth and proliferation under normal conditions, and the lack of mTOR in embryonic murine heart synthesis and cell growth and proliferation under normal conditions, example, Akt activity was upregulated in response to both BAA and DBT treatment. Akt/mTOR signaling promotes protein synthesis and cell growth and proliferation under normal conditions, and the lack of mTOR in embryonic murine heart synthesis and cell growth and proliferation under normal conditions, example, Akt activity was upregulated in response to both BAA and DBT treatment.

In contrast, the activity of the MAPK cascade was upregulated in DBT-treated samples and downregulated in BAA-treated embryos. Evidence from mouse models suggests a complex role for the MAPK pathway in cardiac pathologies. On the one hand, activation of Erk1/2 by the MAPK cascade is implicated in the adaptive cardioprotective response. Its inactivation is associated with a transition from compensatory cardiac hypertrophy to maladaptive hypertrophy and heart failure. On the other hand, Erk1/2 is activated in response to pressure overload and chemotherapeutic treatment and participates in the progression of cardiac pathologies. Thus, activity of the Erk1/2 node in the network may either indicate a compensatory response or implicate the MAPK pathway in cardiotoxic outcomes. In the future, transcriptome analysis at additional time points should help to uncover the dynamics of Erk1/2 activation during exposure to PAHs and in turn help to decipher the role of MAPK signaling in the cardiotoxic effects exerted by PAHs.

BAA treatment resulted in increased Nppa and Nppb activity. During embryonic development, both proteins are expressed in the heart and play a role in cardiac morphogenesis. Under cardiac stress, however, the expression levels of Nppa and Nppb are strongly increased, which makes them reliable markers of heart disease. For example, Nppa is upregulated in the zebrafish model of hypertrophic cardiomyopathy. Additionally, embryonic exposure to PAHs leads to elevated Nppa expression and hypertrophy in adult zebrafish. The cardiotoxicity network analysis predicts that Nppa and Nppb are activated by BAA as early as 48 hpf and can be used as early markers of cardiac stress. Sirtuin 6 (Sirt6) was inferred to be downregulated by BAA. Sirt6 is a stress-responsive protein that plays a role in metabolism, aging, and disease. Reduced Sirt6 expression levels have been reported in the failing human heart as well as in experimental murine models of cardiac hypertrophy. Our results suggest that the cardiac toxicity induced by BAA in zebrafish shares molecular mechanisms with mammalian cardiac hypertrophy.

NO was upregulated in response to DBT treatment. NO is an essential regulator of cardiac contractility and mitochondrial respiration. However, excessive NO levels can result in heart failure and might underlie the cardiac phenotype induced by DBT. Bmp/Smad signaling was activated by both BAA and DBT treatment. Smad signaling is involved in the initiation and progression of many heart diseases and possibly plays a role in PAH-induced cardiotoxicity. However, Bmp/Smad signaling is essential for cardiac development during embryogenesis; therefore, it is possible that the Bmp/Smad activation observed here reflects interference with cardiac development, rather than a specific cardiotoxic effect. Finally, Goodale et al. identified major transcription factors predicted to regulate differentially expressed genes in the data set. Among the transcription factors identified as significant were Jun and Creb1a1. In agreement, both of these nodes are also significant contributors to the network perturbation.

Network Scoring with Sorafenib. In the second data set, Kawabata et al. treated zebrafish larvae with two concentrations of sorafenib. The authors reported that 3 μM sorafenib induced cardiac edema and high mortality in zebrafish larvae, while 1 μM sorafenib did not cause morphological changes, but resulted in decreased ventricular dimension, suggesting impairment of cardiac function. The transcriptome of zebrafish treated with 1 μM sorafenib showed significant perturbation of the cardiotoxicity network, indicating that the network analysis is sensitive to not only anatomical changes but also physiological changes in the heart.

NO, Creb1a1, and Bmp/Smad were inferred to be upregulated upon sorafenib exposure. Additionally, Tomm70a and Opal were upregulated in sorafenib-treated samples. The expression levels of Tomm70a, a translocase of the mitochondrial outer membrane, and Opal, its downstream effector in the mitochondria, show a general decrease in the pathological myocardium of rats and mice. However, this ultimate decline in expression is preceded by a significant increase in Tomm70 protein levels in the first 3 days after initiation of cardiac hypertrophy. Furthermore, zebrafish larvae treated with chlorpyrifos oxon (the active metabolite of the pesticide chlorpyrifos) for 24 h show increased Tomm70a and Opal expression. Similarly, Kawabata et al. conducted the sorafenib exposure experiment for 24 h. Thus, it is possible that early upregulation of Tomm70a and Opal is a compensatory response to a chemical challenge that, under sustained exposure, progresses to downregulation of these genes and eventually results in mitochondrial dysfunction.

Network Scoring with TCE. Lastly, Wirbisky et al. treated zebrafish larvae with TCE, an environmental pollutant that elicits developmental cardiotoxicity in chicks and rats. The cardiotoxicity network was significantly perturbed in response to TCE treatment. The authors reported that TCE-treated larvae displayed mitochondrial dysfunction and decreased angiogenesis. In agreement with the observed phenotype, scoring of the cardiotoxicity network with this data set revealed...
downregulation of mitochondrial Opa1 and NADH dehydrogenase complex I as well as downregulation of vascular-associated Vegfaa and its receptor Kdrl.

Network scoring revealed that TCE treatment induced ROS production and increased MAPK signaling. Mitochondrial dysfunction is widely implicated in cardiotoxicity. Specifically, decreased complex I activity is associated with increased ROS production and apoptosis. Elevated ROS level, in turn, has been implicated in promoting hypertrophic gene expression via the MAPK cascade. Additionally, bovine coronary endothelial cells exposed to TCE show increased ROS levels, decreased NO levels, and decreased proliferation. Accordingly, NO was inferred to be downregulated in TCE-treated larvae. Lastly, Vegf- and Kdrl-dependent angiogenesis is essential for compensatory cardiac response to stress, and inhibition of this process promotes the transition from compensation to heart failure in mice. LN analysis accurately captured the perturbation of these pathways in TCE-treated zebrafish larvae and pinpointed the molecular mechanisms underpinning the observed phenotype.

Limitations and Future Directions. A network model based on scientific literature highly depends on the source material. Curation of unreliable data is likely to have effects on the subsequent scoring results. We mitigated this possibility in the following ways: all the papers were curated manually and judged to be of sufficient quality; gap-filling during network construction encouraged re-evaluation of the construction at each point; although the edges in the network have evidence from at least one paper, many edges have multiple lines of evidence, thus reducing sensitivity to single results; network perturbation measures effects on the network as a whole, thus also reducing sensitivity to single edges that may prove to be unreliable.

Of the five cardiotoxins evaluated in this study, we were able to correctly identify four. The results suggest that the time at which gene expression measurements are made is very important for using the presented approach in an optimal way. We did not detect cardiotoxicity at 24 hpf with any of the chemicals tested. This indicates that our approach is suitable for larval stages but, with the current cardiotoxicity network, should not be used for early embryonic stages. At the moment, there is no standardized method for collecting transcriptomic data for toxicological assessment. Experimental designs vary in terms of chemical concentration in relation to phenotypic (EC) or lethal (LC) effects, presence or absence of solvents, embryo age at the start of exposure, exposure duration, presence or absence of the chorion, and light and temperature conditions. The transcript layer of our causal network, therefore, may not always reflect the experimental conditions of the query transcriptome. Our transcript layer includes data from many different studies that measured gene expression at different times after chemical exposure. For example, transcripts measured 8 h after activation of a node in the network might be mapped to that node. Scoring of the network with a transcriptome collected 48 h after this node was active might not identify this activation because of the time discrepancy. Indeed, the cardiotoxic effects of PAHs are widely thought to be initiated by activation of the aryl hydrocarbon receptor Ahr2. However, in the present study, network scoring did not identify Ahr2 as a LN. With the growing popularity of the zebrafish as a model organism, and with more standardized transcriptomic data becoming available, we intend to continuously update the current version of the model (v1.0) on causalbionet.com to improve its accuracy. To address the lack of standardized publicly available transcriptomic data sets, we are conducting a chemical screen in accordance with the OECD fish embryo toxicity test to compare network scoring with transcriptomic data and apical end points. Chemicals with known cardiotoxic effects will be used to validate the network. Conversely, network scoring results will be used to investigate chemicals with unknown effects. We envisage such an approach to be useful to study not only environmental chemicals but also potential drugs for on or off-target effects, food additives, and chemical mixtures.

Some nodes and even whole pathways in the cardiotoxicity network play roles outside of the heart and may be detected as unspecific events. To mitigate this, the NPA measures the impact on the network as a whole and not on isolated pathways. Nevertheless, it is currently difficult to say how specific our approach is to cardiac toxicity. We expect this to be clearer in the future as more data sets are scored with the network.

The noncomputer-readable language used in scientific publications necessitates extensive manual curation of scientific findings before they are ready as input for computational approaches. The semiautomated approach used in BELIEF addresses this to an extent. However, human contribution remains a necessary and thus a rate-limiting step. It is tempting to envisage the future emergence of a central repository of BEL statements with associated literature and annotations, verified by researchers as a collective, akin to resources like Wikipedia. One immediately available approach for alleviating this congestion could be for authors of peer-reviewed manuscripts to include BEL statements supported by evidence in their publications. This would greatly facilitate text-mining and network model building.

CONCLUDING REMARKS

In summary, we have presented a method for cardiotoxicological assessment of chemicals based on transcriptomic data. This method helps extract data of interest from the whole transcriptome. In this regard, the cardiac network can be used to filter out unrelated information and reveal chemical toxicity specifically in the heart. The method includes a measure of perturbation of the network (NPA) and an analysis of the main events and pathways responsible for the perturbation. This represents a big step forward compared with conventional transcriptomics analysis. Although conventional methods are very useful for finding novel molecular mechanisms of chemical-induced toxicity, they do not easily allow us to connect transcriptomic changes to the phenotype of interest. We have shown that the NPA score matches the observed phenotypes even when transcriptomic samples are taken several days before the phenotype is observed, but after major developmental events have concluded. The reduction of high-dimensional transcriptomic data into a single NPA allows comparison between several conditions such as disease progression stages, dose responses, time courses, and treatments. An additional advantage is that only one experimental assay (transcriptomics analysis) is needed to infer the main toxic molecular events, circumventing the need for different assays with different levels of complexity. However, the accuracy here depends on the quality of the transcriptomic data sets in the transcript layer of the network and should, therefore, be interpreted with some caution, especially in cases when the transcript layer data sets and the data set used for evaluation come from different experimental designs. While further testing of this approach is necessary in both cardiotoxicological assessment and other types
of toxicity, we believe that the developed network and scoring approach will be useful for both human and ecotoxicological chemical screening.

**ASSOCIATED CONTENT**

1. Supporting Information
The Supporting Information is available free of charge at https://pubs.acs.org/doi/10.1021/acs.chemrestox.0c00095.

Curation workflow (PDF)
Publications curated to create the network (XLSX)

**AUTHOR INFORMATION**

*Corresponding Authors*
Marja Talikka — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland; Email: Marja.talikka@pmi.com

Roman Li — Swiss Federal Institute of Aquatic Science and Technology, Eawag, CH-8600 Dübendorf, Switzerland; PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland; orcid.org/0000-0002-0290-9619; Email: roman.li@eawag.ch

*Authors*
Anze Zupanic — Swiss Federal Institute of Aquatic Science and Technology, Eawag, CH-8600 Dübendorf, Switzerland; orcid.org/0000-0003-3303-9086

Vincenzo Belcastro — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland

Sumit Madan — Fraunhofer Institute for Algorithms and Scientific Computing, Schloss Birlinghoven, Sankt Augustin 53754, Germany

Jens Dörpinghaus — Fraunhofer Institute for Algorithms and Scientific Computing, Schloss Birlinghoven, Sankt Augustin 53754, Germany

Colette vom Berg — Swiss Federal Institute of Aquatic Science and Technology, Eawag, CH-8600 Dübendorf, Switzerland; orcid.org/0000-0002-8870-9799

Justyna Szostak — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland; orcid.org/0000-0002-6233-612X

Florian Martin — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland

Manuel C. Peitsch — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland

Julia Hoeng — PMI R&D, Philip Morris Products S.A., CH-2000 Neuchâtel, Switzerland

Complete contact information is available at: https://pubs.acs.org/10.1021/acs.chemrestox.0c00095

**Funding**

**Notes**
The authors declare no competing financial interest.

**ABBREVIATIONS**

BEL, biological expression language
BELIEF, BEL information extraction workflow
NPA, network perturbation amplitude
LN, leading node
OECD, organization for economic co-operation and development

AOP, adverse outcome pathway
ZFIn, zebrafish information network
PAHs, polycyclic aromatic hydrocarbons
BAA, benz[a]anthracene
DBT, dibenzothiophene
PYR, pyrene
TCE, trichloroethylene

**REFERENCES**


Chemical Research in Toxicology


