



OPEN

Ontological approach to the knowledge systematization of a toxic process and toxic course representation framework for early drug risk management

Yuki Yamagata^{1,2}✉ & Hiroshi Yamada¹✉

Various types of drug toxicity can halt the development of a drug. Because drugs are xenobiotics, they inherently have the potential to cause injury. Clarifying the mechanisms of toxicity to evaluate and manage drug safety during drug development is extremely important. However, toxicity mechanisms, especially hepatotoxic mechanisms, are very complex. The significant exposure of liver cells to drugs can cause dysfunction, cell injury, and organ failure in the liver. To clarify potential risks in drug safety management, it is necessary to systematize knowledge from a consistent viewpoint. In this study, we adopt an ontological approach. Ontology provides a controlled vocabulary for sharing and reusing of various data with a computer-friendly manner. We focus on toxic processes, especially hepatotoxic processes, and construct the toxic process ontology (TXPO). The TXPO systematizes knowledge concerning hepatotoxic courses with consistency and no ambiguity. In our application study, we developed a toxic process interpretable knowledge system (TOXPiLOT) to bridge the gaps between basic science and medicine for drug safety management. Using semantic web technology, TOXPiLOT supports the interpretation of toxicity mechanisms and provides visualizations of toxic courses with useful information based on ontology. Our system will contribute to various applications for drug safety evaluation and management.

Liver toxicity is a major cause of attrition in drug development and the withdrawal of drug products. Drugs inherently have the potential to cause injury because they are xenobiotics. However, it is difficult to clarify the mechanisms of toxicity because of their complexity. The liver is the main organ where chemicals are metabolized and eventually excreted. Therefore, significant exposure of liver cells to drug and drug candidates can cause liver dysfunction, cell injury, and liver organ failure¹. For drug risk management, it is essential to systematize the necessary knowledge from a consistent viewpoint.

In this study, we adopt an ontological approach. Originally, ontology was a branch of philosophy, and the term means 'existence'². Recently, information science has focused on ontology, because ontology provides a controlled vocabulary for sharing and reusing various data in a computer-friendly manner³. There are many biomedical ontologies in ontology repository sites such as the National Center for Biomedical Ontology (NCBO) BioPortal (<https://bioportal.bioontology.org>). However, none of these ontologies organize toxicity mechanisms. To systematize knowledge concerning toxic mechanisms without ambiguity, we constructed a toxic process ontology (TXPO) focusing on hepatotoxic processes.

To analyse toxicity mechanisms for drug safety management, a unified viewpoint is needed to cope with the range of granularities in the human body. Toxicants can disrupt unexpected functions at the molecular, cellular, and/or tissue levels¹. There are many biomedical pathway databases, including the Kyoto Encyclopedia of Genes and Genomes (KEGG)⁴, WikiPathways⁵, and Reactome⁶. Because these databases deal with many pathways, one

¹Toxicogenomics Informatics Project, National Institute of Biomedical Innovation, Health and Nutrition, 7-6-8 Saito-Asagi, Ibaraki, Osaka 567-0085, Japan. ²Laboratory for Developmental Dynamics, RIKEN Center for Biosystems Dynamics Research, 2-2-3 Minatogijima-minamimachi, Chuo-ku, Kobe 650-0047, Japan. ✉email: yuki.yamagata@riken.jp; h-yamada@nibiohn.go.jp

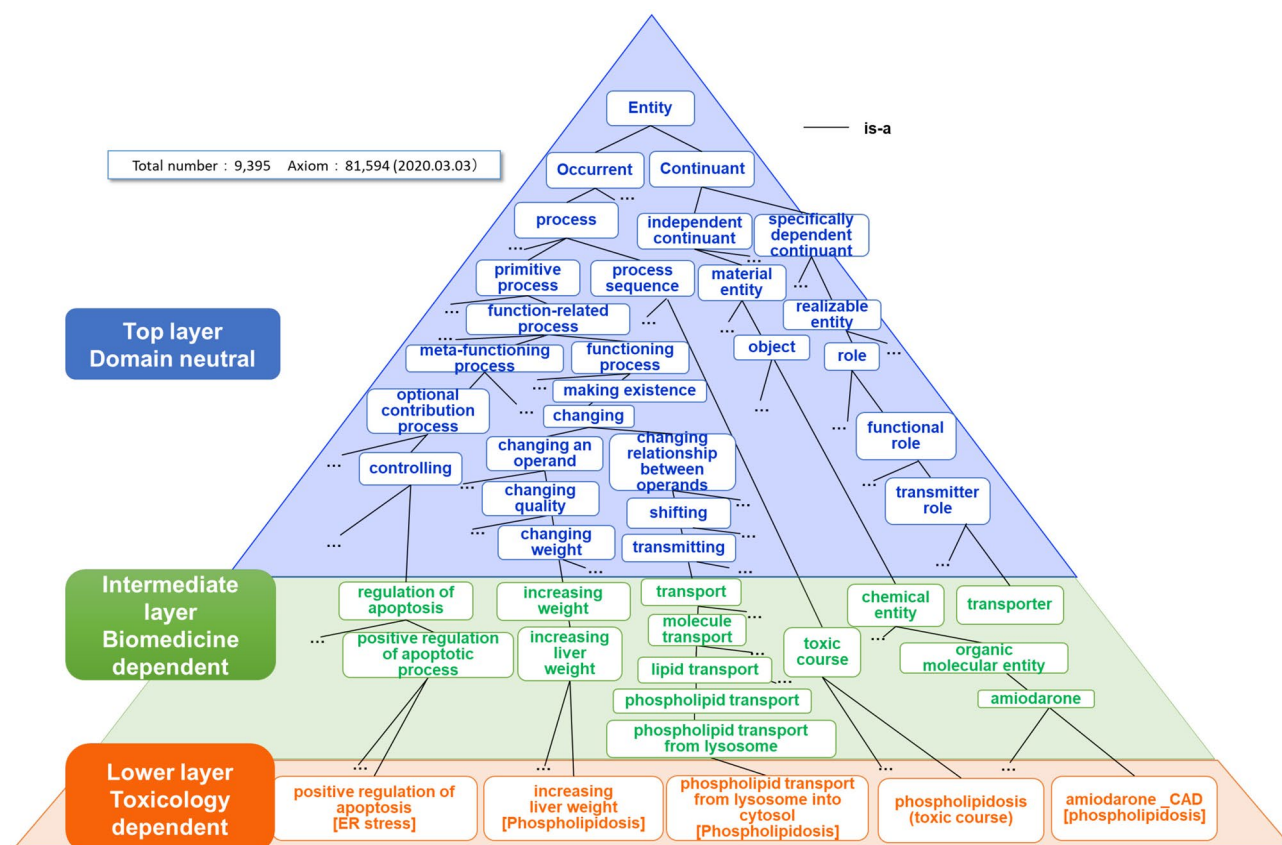


Figure 1. Overview of the ToXic Process Ontology (TXPO), which contains a three-layer *is-a* hierarchy: the top layer contains general terms, mostly from the Basic Formal Ontology (BFO). The intermediate and lower layers contain biomedical terms and more granular toxicology terms, respectively.

might conclude that they also explain toxicity mechanisms. However, most of them are based on molecule–molecule interactions. Such molecular-centred approaches do not cover cell- or organ-level granularity. In this study, we focus on functioning processes and introduce a functional decomposition approach. We capture a biological system and its parts as sub-systems to describe biological defence functions from a consistent viewpoint across granularities.

Screening in drug discovery requires strategies to find and reduce drug risk before a toxic manifestation. An adverse outcome pathway (AOP)⁷ constructs a representation of biological events leading to adverse effects, and it provides a coarse granularity of events. However, an AOP mainly focuses on measurable changes (key events). In this study, we modelled a representation framework that describes toxic courses from latent to toxic manifestations. We represent a toxic course as causal relationships between toxic processes and introduce a model of the imbalance between biological defence functioning processes and toxic action processes. As a case study, we describe phospholipidosis. Because the imbalance model can represent the adaptation stage, it will contribute to early drug risk management by enabling an appropriate subpopulation of patients to be identified.

Results

Development of the TXPO for toxic mechanism knowledge systematization. *Outline of the TXPO.* The TXPO is a three-layer model organized in an *is-a* hierarchy from general terms to specialized toxicological terms (Fig. 1). The top layer is domain-independent and provides general terms from the upper-level Basic Formal Ontology (BFO)². BFO supports generic categories and relations based on a philosophical orientation. Accordingly, we constructed our ontology using the inheritance of the intrinsic nature of an entity in a consistent manner. All entities of the TXPO are classified as *continuant* or *occurent*. The term *continuant* refers to an entity that persists, endures, or continues to exist through time while maintaining its identity and includes *objects*, *roles*, and *qualities*. An *object* is an independent continuant, such as a thing. *Roles* and *qualities* are dependent *continuants* that can only exist depending on something else. *Occurrent* includes entities that unfold over time, such as *processes*.

The intermediate layer consists of biomedical entities. These terms include anatomic structures from Uberanatomy ontology (UBERON)⁸, cells from the Cell Ontology⁹, organisms from the NCBI Taxonomy¹⁰, compounds from Chemical Entities of Biological Interest (ChEBI)¹¹, biological processes and cellular components from the Gene Ontology (GO)¹², qualities from the Phenotype And Trait Ontology (PATO)¹³, some molecule

families from INOH¹⁴, genes from the Ontology of Genes and Genomes (OGG)¹⁵, and diseases from the Disease Ontology¹⁶. The size of each ontology is vast, and many parts are unrelated to toxic mechanisms. Therefore, only those terms relevant to the toxic courses were manually imported.

The lower layer consists of entities specific to toxicology.

Process in the TXPO. *Process* is a central category in the TXPO. To elucidate a toxicity mechanism adequately, we provide two sub-categories: *primitive process* (a process that is a single unit) and *process sequence* (a series of processes, which includes pathways and toxic courses).

1. Functioning processes

The function of many biological defence processes is to protect organisms from toxicity-associated injury. Therefore, we focus on functioning processes. Functioning processes in organisms vary in granularity, from the molecular level to the organelle, cell, tissue, and organ level. To define functioning processes consistently, we refer to functional ontology¹⁷. Functional ontology defines general functions based on changes in the state of the input–output relationship among physical things and models functional knowledge. The subclasses of *functioning process* in the TXPO top layer are imported and reused from functional ontology (Fig. 2). In detail, a *functioning process* can be mainly categorized into *receiving*, *making existence*, and *generating* categories. The *making existence* category can be further subdivided into *changing an operand* and *changing relationship between operands* classifications. *Changing an operand* includes changing qualities such as concentration, volume, or weight. Examples of subtypes of *changing relationship between operands* are *sifting* and *separating*. Subtypes of *sifting* include *transmitting*, and subtypes of *separating* include *decomposing*, *splitting*, and *detaching*. By specializing the entities in the top layer, we define original terms in the intermediate and lower layers in the TXPO. The intermediate layer is biomedical-domain dependent. For instance, a lower level of *transmitting* includes biological transport processes such as lipid transport. *Decomposing* includes biochemical degradation; *splitting* includes cell division; and *detaching* includes complex dissociation. Some entities can be mapped to GO biological processes, as some of GO terms can be interpreted as functional processes common to biomedicine.

The lower layer is toxicology-domain dependent. In this study, we define a ‘toxic process’ as a process that constitutes a specific toxic course. For example, to specialize the ‘phospholipid transport (GO: 0015914)’ in GO, we define an ‘phospholipid transport from lysosome into cytosol [Phospholipidosis]’ that constitutes a course of phospholipidosis.

One difficulty in defining a toxic process is that protective responses to drugs can also injure cells. Hence, to understand the toxicity mechanisms appropriately, we also regard a process that functions as a biological defence in the toxic course as a ‘toxic process’.

The development of the toxicity-dependent process subtree was based on the ‘low-hanging fruit’ policy. Terms were extracted and manually annotated from toxicology-related textbooks and articles.

In addition to the function-execution process, the TXPO defines *meta-functioning processes* as function-related processes specific to other functions and includes *controlling*, for example. Subtypes of *controlling* include the regulation of apoptosis and cell cycle control.

2. Decomposition of functioning

The TXPO specifies a functioning process based on a function decomposition framework. In an ontological engineering approach, a device (system) consists of sub-devices (sub-systems). In a function decomposition tree, the overall function of a system is achieved by a sequence of sub-functions of the sub-systems. Because biological functions can be considered specializations of systemic functions¹⁸, we attempted here to clarify the functioning process of biological structures for each granularity based on the whole–part relationship (part-of/has-part relationship). At the cell level, we regard a cell as a system and cell components such as organelles as system parts. Figure 3 shows an example describing how the cell system decreases phospholipids to maintain phospholipid homeostasis. Under normal conditions, the cell system maintains the phospholipid level by phospholipid import/export, phospholipid biosynthesis (anabolism)/catabolism, and similar processes. In the course of phospholipidosis, it is known that cationic amphiphilic drugs (CADs) bind to the lysosomal membrane, which triggers a negative regulation of phospholipid degradation, resulting in the accumulation of phospholipids in lysosomes. Therefore, the cell system performs a ‘decreasing phospholipid’ process as a biological defence function (enlarged part of Fig. 3). In the cell system, the organelle performs a sub-function as a sub-part. For example, ‘phospholipid catabolic process’ and ‘phospholipid transport from lysosome into cytosol’ work in lysosomes, and ‘lysosomal enzyme transport’ executes in the late endosome. Moreover, in the nucleus, a function that promotes degradation, ‘phospholipase gene expression’ is performed. Accordingly, the cell system reduces phospholipids by performing sub-functioning processes of the system parts.

However, when the accumulation of phospholipid becomes more severe, autophagy works to remove phospholipid mass in lysosomes, and apoptosis also plays a defensive role in removing phospholipid-accumulated cells in the liver. In this way, the functional decomposition approach can specify a variety of defence functions using whole-of-part relationships of a biological system within a unified representation framework. Furthermore, the function decomposition tree reveals that dysfunction in one process affects other processes from the bottom up. Therefore, this approach can help predict the progress of toxicity and drug risk management so that an alternative path can be found.

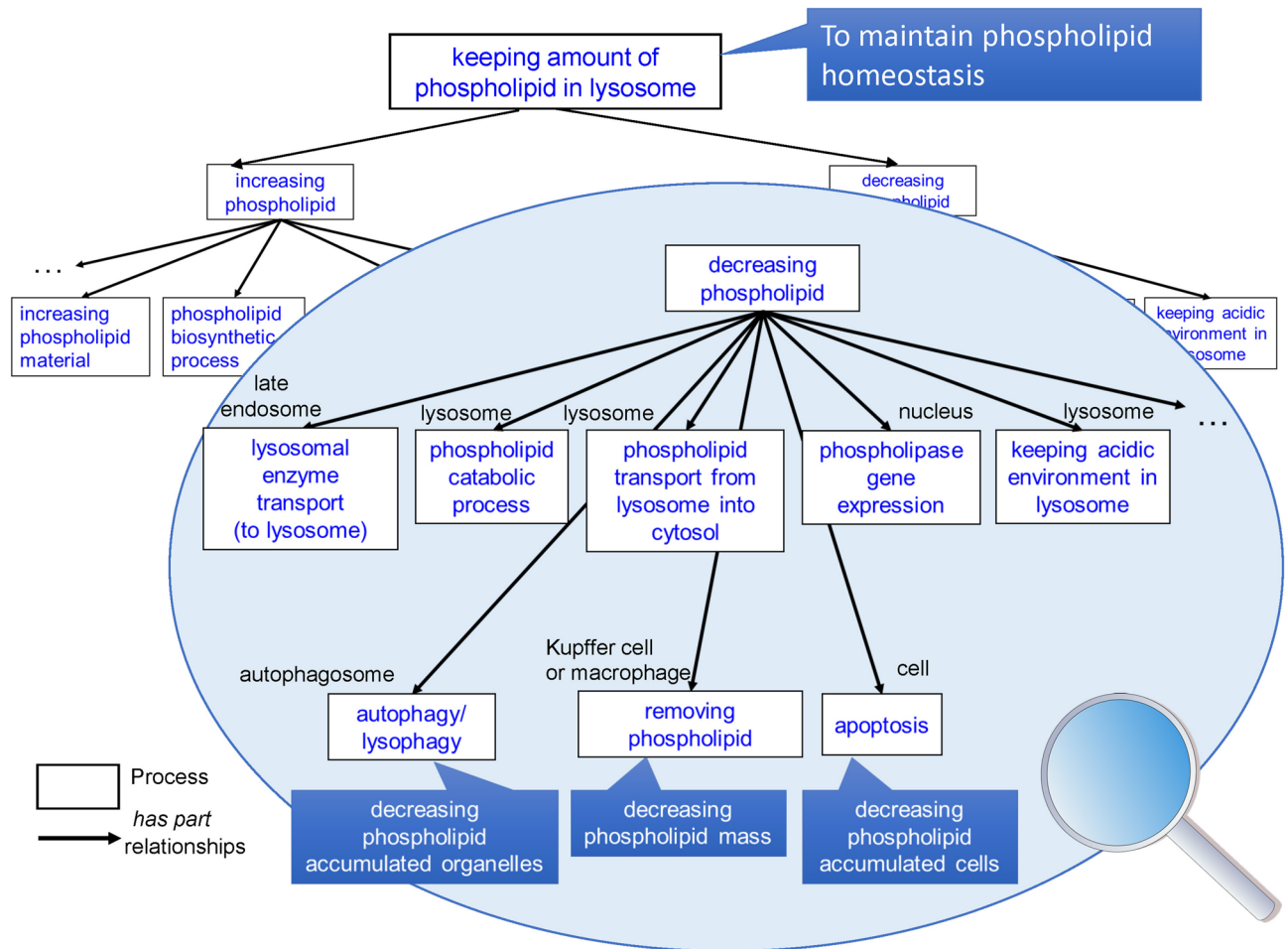


Figure 3. Example of a functional decomposition of ‘decreasing phospholipid’.

3. Toxic course definition

In the domain of toxicology, understanding a toxicity mechanism is crucial for drug safety management. Toxicity mechanisms are generally explained in terms of multiple processes such as toxicant delivery, biological defence process, cellular dysfunction/dysregulation, and cell death¹. Therefore, to clarify the toxicity mechanism, we define a toxic course as a series of process in an organism that manifests toxicities and that are not part of the life of the organism. Processes in one toxic course are represented by the causal relationships between processes. Furthermore, the processes of a parent toxic course are inherited by the child toxic courses (Supplementary Information 1). Figure 4 illustrates an example of the causal relationships of phospholipidosis and the child toxic course sphingomyelin disorder. Drugs like CADs are known to negatively regulate phospholipid degradation in lysosomes¹⁹. In this study, phospholipidosis includes the following causal (has-result) relationships, indicated using ‘→’ (Fig. 4a):

Compound accumulation in lysosome

- → negative regulation of phospholipid degradation
- → hypofunction of phospholipid degradation
- → phospholipid metabolism imbalance
- → phospholipid accumulation in lysosome

All instances of sphingomyelin disorder are also instances of the parent toxic course, phospholipidosis. They include the causal relationships of the processes in phospholipidosis. Moreover, by specializing phospholipid into sphingomyelin, the course of sphingomyelin disorder holds the following causal relationships (Fig. 4b):

Compound accumulation in lysosomes

- → negative regulation of sphingomyelin catabolic process
- → hypofunction of sphingomyelin degradation
- → sphingomyelin metabolism imbalance
- → sphingomyelin accumulation in lysosome

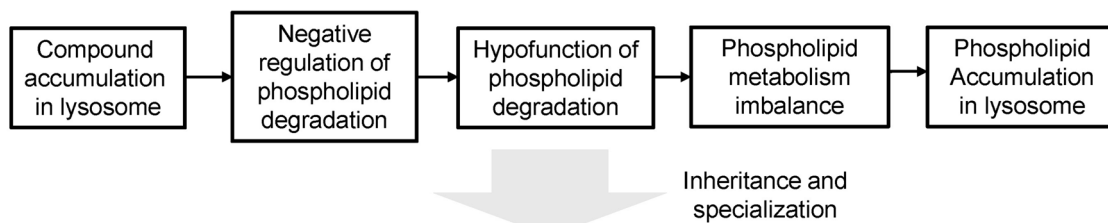
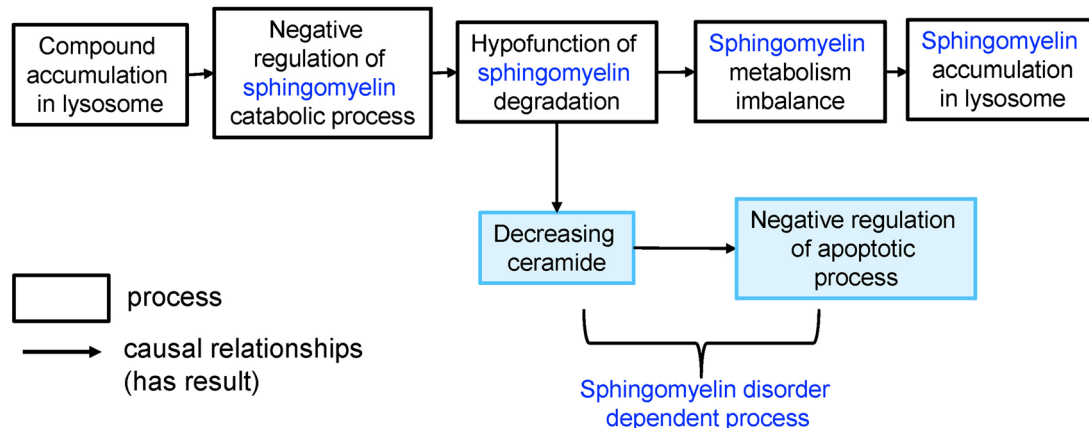
(a) Phospholipidosis (parent) URI: http://purl.obolibrary.org/obo/TXPO_0002215(b) Sphingomyelin disorder (child) URI: http://purl.obolibrary.org/obo/TXPO_0004157

Figure 4. Examples of causal relationship representations of toxic courses: (a) phospholipidosis, (b) sphingomyelin disorder. Blue text indicates the specialization from phospholipid to sphingomyelin; blue boxes show additional sphingomyelin-dependent processes.

In TXPO, a parent course can also be specialized into a child course by adding specific processes. For example, ‘hypofunction of sphingomyelin degradation’ can cause the following processes specific to the course of sphingomyelin disorder:

Hypofunction of sphingomyelin degradation

- → decreasing ceramide
- → negative regulation of apoptotic process

The toxic course currently contains more than ten courses on the following themes: ER stress, glutathione depletion, phospholipidosis, lipidosis, mitochondrial damage, ground glass appearance, eosinophilic granular degeneration, hepatocarcinogenesis, cell death, and increasing liver weight.

Role. In general, a molecule plays multiple roles in the body. Therefore, our goal was to explicate the roles of molecules participating in specific processes in the toxic course. For example, PLA2G15 (phospholipase A2 group XV) can act as a ‘positive regulator of immune response’ in the inflammatory response in severe phospholipidosis. As for drugs, TXPO makes their roles explicit in a specific toxic process. For example, amiodarone plays the roles of a ‘cationic amphiphilic drug’ and ‘competitive inhibitor of phospholipase’ and participates in the negative regulation of phospholipid degradation in phospholipidosis. Amiodarone also takes on a ‘mitochondrial respiratory-chain inhibitor’ role in the negative regulation of the respiratory electron transport chain in lipidosis.

Relationships among entities. As of March 3, 2020, 9,395 entities have been defined in the TXPO. Figure 5 shows an example of the relationships among the terms defined in phospholipidosis.

Toxic course map as a representation framework for early risk management of drugs. To address drug safety management from an early stage, it is important to represent the processes before the onset of toxicity for computer analysis. Cells constantly adapt to physiological demand to maintain a homeostatic steady state²⁰, and adaptation can be thought as a process to maintain homeostasis against functional demand. In this study, we introduce a demand and supply imbalance model²¹ into the toxic course representation. In this model, supply indicates the functioning processes associated with biological defence and demand refers to toxic activity. The basic units are as follows (Supplementary Information 2):

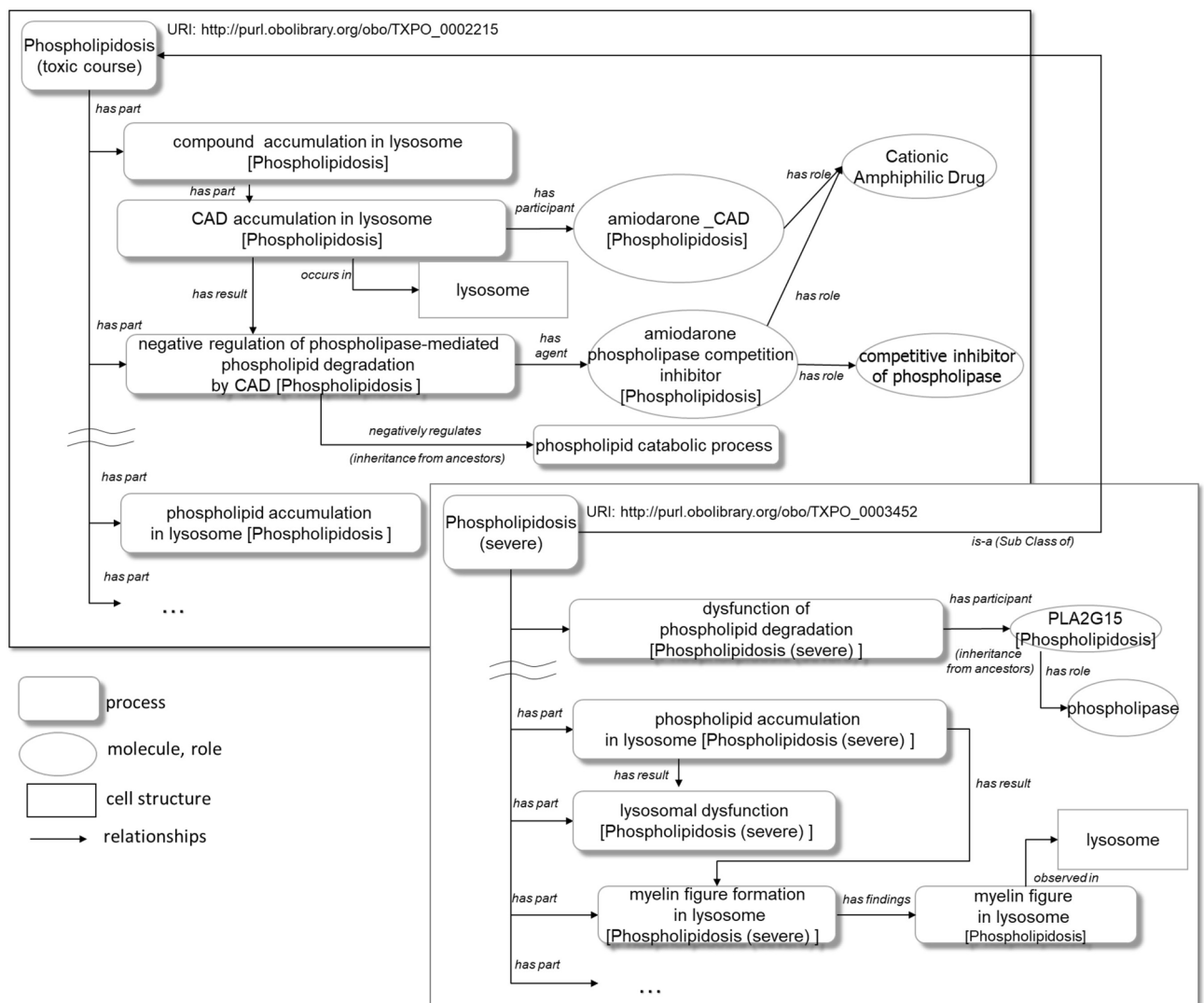


Figure 5. Examples of TXPO relationships. Rounded rectangles indicate toxic courses and processes; ellipses indicate molecules and their roles. Boxes represent biological structures, e.g. lysosome. Not all processes are shown. In the phospholipidosis toxic course, for instance, CAD accumulation in lysosome process *occurs in* lysosome and amiodarone *has role* cationic amphiphilic drug.

1. a functioning process (supply) as a biological defence for maintaining homeostasis;
2. a functional demand process (demand) as toxic activity;
3. balance/imbalance between toxic activity and defence processes;
4. outcome of the organelles, cells, or tissues of the organ exhibiting toxicity manifestations.

The level of functioning performance can change according to changes in demand; however, if demand exceeds performance, an imbalance and outcomes that manifest toxicity occur.

To understand the precise mechanisms in idiosyncratic toxicity, the levels of functional performance must be considered. Here, we introduce the following levels: ‘very low’, ‘low’, ‘medium’, ‘high’, and ‘very high’. Since cells normally maintain homeostasis in which the milieu is maintained within a narrow range²², we define this level of functioning as ‘medium’. In the body’s system, if the functional demand increases to a high level, the defence function also performs at a high level to adapt, and a new homeostasis is obtained. However, under severe stimulus conditions, e.g. drug exposure, a ‘very high’ toxic action results in an imbalance and can lead to irreversible cell injury and cell death. When a ‘very high’ level of defence function exceeds demand, it can also lead to serious damage such as liver fibrosis (details are given in Supplementary Information 2).

Imbalance model in phospholipidosis. Here, we applied the imbalance model to an example of a toxic course, phospholipidosis. In phospholipidosis, the basic units are (1) increasing the phospholipid process as functional demand to decrease phospholipid, (2) decreasing the phospholipid process as a defence function, (3) phospholipid homeostasis balance/imbalance between functional demand and function, and (4) the ‘phospholipid accumulation’ process as an outcome. In the drug exposure situation in phospholipidosis, it is known that CADs

affect biological functioning. CADs bind to lysosomal membranes and form a complex with phospholipids, which negatively regulates phospholipid degradation¹⁹. Under normal conditions, the cell system can maintain a balance between ‘phospholipid biosynthetic process’ and ‘phospholipid degradation’ at a ‘medium’ level. However, when the CADs are not excreted from the body, the defence system cannot perform at a ‘high level’, i.e. there is hypofunction of phospholipid degradation. Although no outcomes occur, the cellular system is in the so-called latent stage (Fig. 6a) Here, if the functional demand is increased by additional causes such as ‘positive regulation of phospholipid biosynthetic process’ caused by an ‘increase in fatty acid inflow into hepatocyte’, then an imbalance can occur. As a result, ‘phospholipid accumulation in lysosome’ manifests as an outcome, i.e. there is a lack of adaptation (Fig. 6b) For example, obesity causes an ‘increase in fatty acid inflow into hepatocyte’, which can promote phospholipid biosynthesis (Fig. 6c). Therefore, obese patients with chronic CAD use could be disposed to have a phospholipid homeostasis imbalance.

In Niemann–Pick disease type A, because of a ‘sphingomyelinase gene mutation’, ‘dysfunction of sphingomyelin degradation’ can occur, which decreases the defence level to ‘very low’ (Fig. 6d). Therefore, an imbalance always occurs regardless of the existence of CADs, and sphingomyelin accumulation is observed as an outcome. Several drugs such as tricyclic antidepressants (e.g. desipramine and imipramine) have been shown to induce functional loss of acid sphingomyelinase activity *in vivo*²³. Desipramine protects hepatocellular apoptosis via the inhibition of ceramide channel formation²⁴, so chronic use of imipramine might induce the accumulation of sphingomyelin, as in Niemann–Pick disease, and inhibit ceramide production.

In summary, there are various patterns of imbalance that can occur in a toxic course. In this study, we focused on hepatotoxicity. However, toxic expression varies depending on the organ function. For instance, in phospholipidosis in the lung, the function of pulmonary surfactant may be impaired, and phospholipidosis in the heart can lead to a life-threatening disorder such as QT prolongation due to cell hypertrophy caused by phospholipid accumulation. In future, we plan to analyse toxic courses in other organs.

Application: development of the TOXPILOT knowledge system. One of the advantages of an ontology is it can assist navigation to appropriate knowledge, which is useful for various applications in drug safety evaluation and management. In this study, we developed a toxic process interpretable support knowledge system called TOXPILOT (<https://toxipilot.nibiohn.go.jp>). The TOXPILOT system provides useful information to help users acquire knowledge concerning toxic process based on the TXPO, which helps bridge the gaps between basic science and medicine and accelerate knowledge sharing across disciplines for drug safety management. Using semantic web technology, TOXPILOT supports the interpretation of toxicity mechanisms and visualizes toxic courses with useful information based on ontology. Retrospective and forward analyses from one common finding demonstrate how the processes of phospholipidosis and non-alcoholic steato-hepatitis (NASH) can be compared.

1. Toxic course map

TOXPILOT provides toxic course maps that can visualize the causal relationships of toxic processes, as described in the previous section (Fig. 7a). In a toxic course map, a toxic course is shown as a graph consisting of nodes representing processes and links representing their causal relationships. This map can also show the molecules that participate in the process and pathological findings.

Our preliminary data show that the maps enable us to visualize marker genes for liver toxicity predicted by support vector machines. We can hence identify genes that are predicted to participate in common processes in ER stress based on rat *in vivo* and human *in vitro* analyses^{25,26}. Thus, this toxic course map facilitates evaluation and extrapolation to humans for translational research.

TOXPILOT also provides process maps based on a functional decomposition approach. These maps enable the visualization of pathologic findings associated with a process.

2. Support for an *in vitro* screening system

Next, we examined whether TOXPILOT can support the prediction of toxic marker genes in an *in vitro* screening system using large-scale DNA microarray gene expression analysis for phospholipidosis. We selected one article on the human *in vitro* gene prediction of phospholipidosis²⁷ and annotated the marker genes predicted in the article with respect to related processes, molecule role, and causal relationships. Figure 7b shows the results. For example, for ASAH1 (N-acylsphingosine amidohydrolase 1) in the sphingomyelin disorder, which is a subtype of phospholipidosis, our map shows that inactivated ASAH1 participates in the ‘hypofunction of ceramide degradation’ process. ASAH1 encodes the enzyme, acid ceramidase. In sphingomyelin metabolism, sphingomyelin degradation produces ceramides, and acid ceramidase catalyses ceramide degradation. However, in the sphingomyelin disorder, ‘hypofunction of sphingomyelin degradation’ can cause ‘decreasing ceramide’, which might change acid ceramidase coding gene ASAH1 expression, and ‘hypofunction of ceramide degradation’ can occur. TOXPILOT also provides access to a wealth of information about each node. For instance, ASAH1 plays a role in the ceramide signalling pathway and other processes such as inducing apoptosis for tumour cell death.

Hence, for *in vitro* screen analysis, it is possible for this system to show marker genes and their roles in related processes according to the mechanisms. In general, genes are known to play multiple roles in the body, and TOXPILOT shows the role of each gene in the specific process according to the mechanism, which contributes to drug safety evaluation.

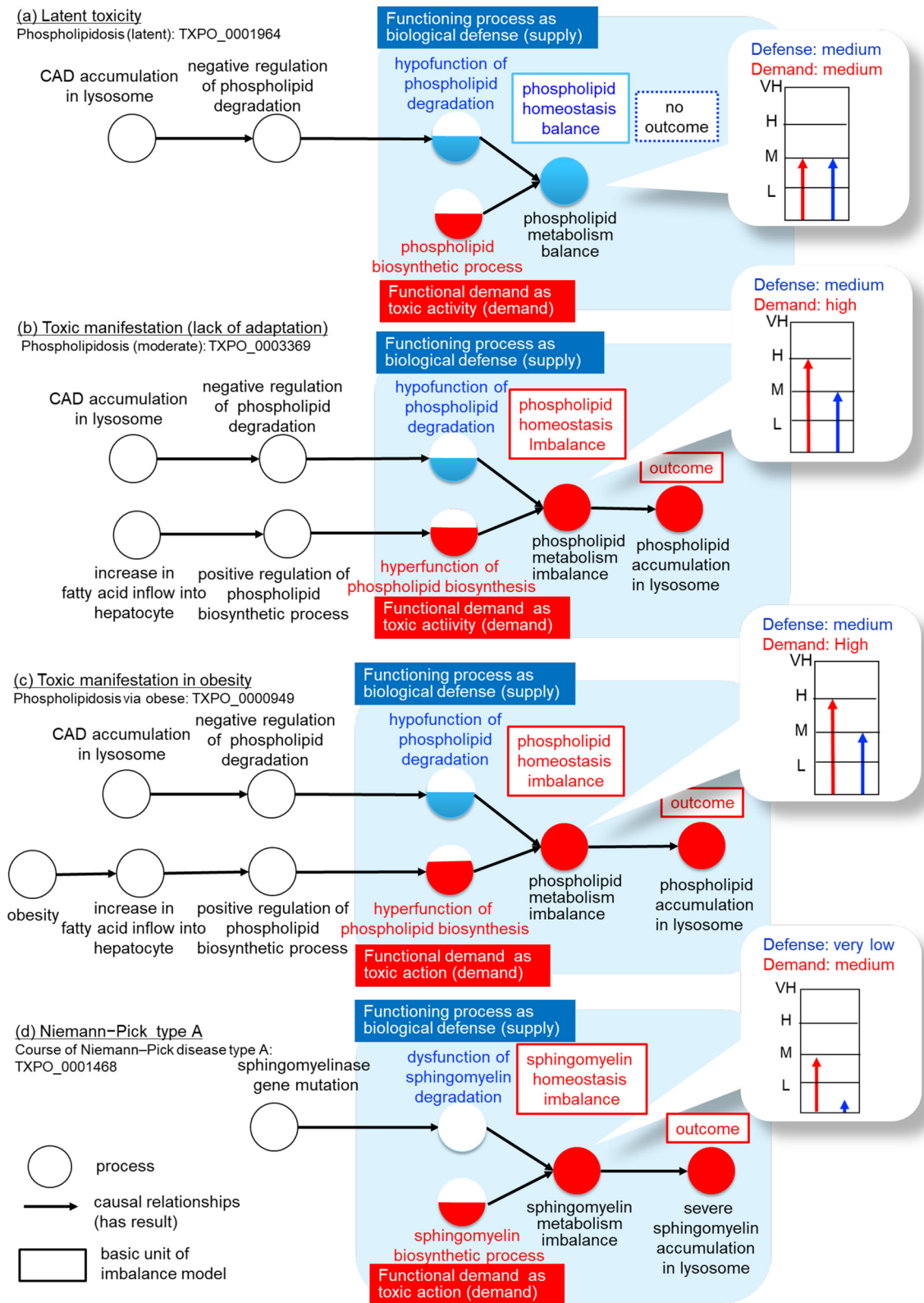


Figure 6. Imbalance model for phospholipidosis. Circles indicate processes; links between nodes indicate causal relationships (has result relationships). Boxes represent the basic units of the model: a functioning process (supply) as a biological defence, a functional demand process (demand) as toxic activity, balance/imbalance, and outcome. Not all processes are shown. **(a)** Latent toxicity (Phospholipidosis (latent): http://purl.obolibrary.org/obo/TXPO_0001964): although balance is maintained, the body system cannot adapt when CADs are present. **(b)** Toxicity manifestation (Phospholipidosis (moderate): http://purl.obolibrary.org/obo/TXPO_0003369): While CAD accumulation causes the negative regulation of degradation, other causes, e.g. positive regulation of phospholipid biosynthesis, cause hyperfunction of phospholipid biosynthesis, causing imbalance and phospholipid accumulation as an outcome. **(c)** Toxic manifestation in obesity (Phospholipidosis via obese: http://purl.obolibrary.org/obo/TXPO_0000949). **(d)** Niemann–Pick type A (course of Niemann–Pick disease type A: http://purl.obolibrary.org/obo/TXPO_0001468).

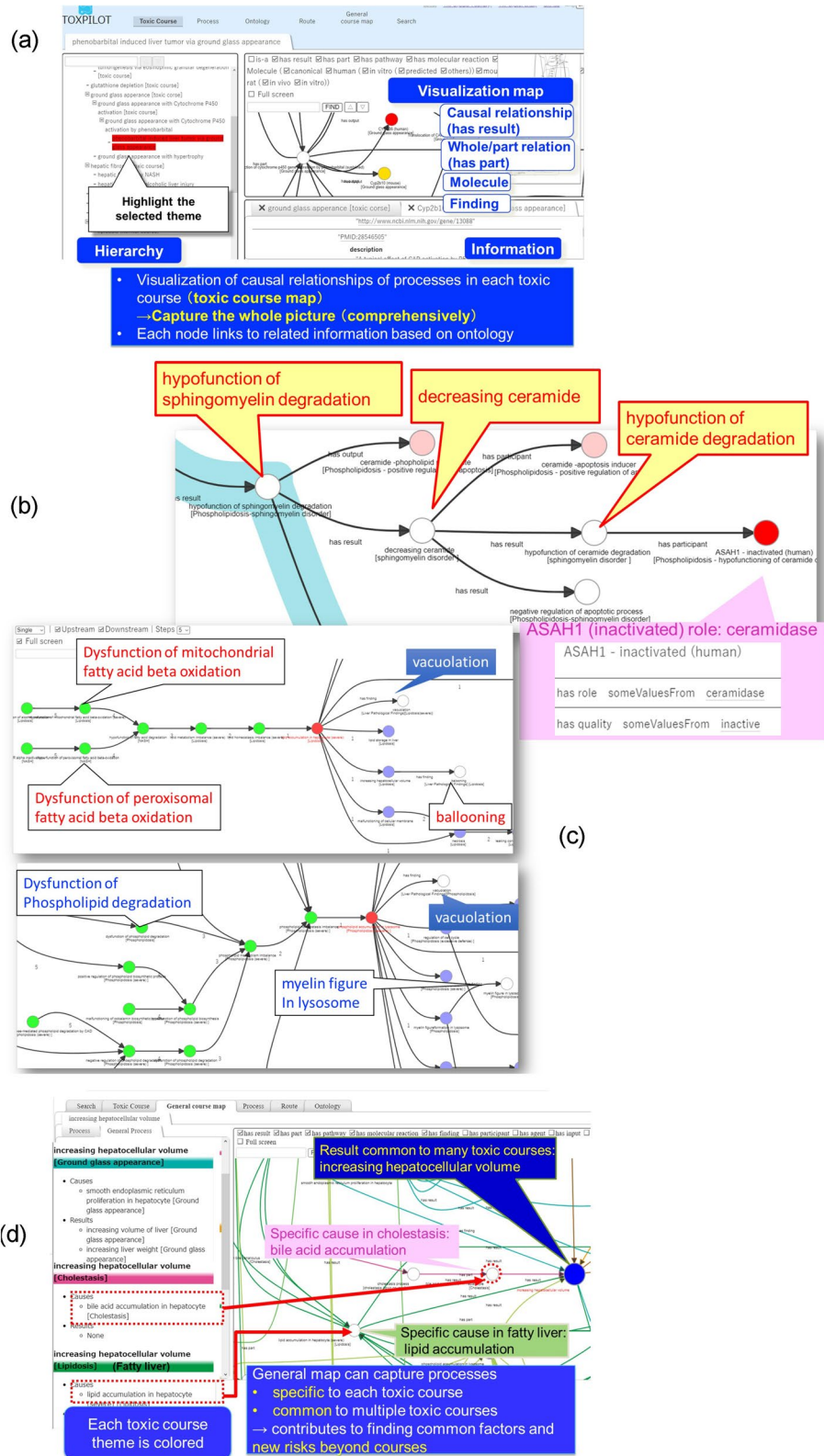


Figure 7. Toxic process interpretable knowledge system (TOXPILOT) **(a)** The toxic course map visualizes a toxic course as causal relationships between processes. **(b)** In vitro marker genes visualized in the phospholipidosis map. **(c)** Route search provides the upstream or downstream paths of the selected process. (Upper) Route search from vacuolation in phospholipidosis. (Lower) Route search from vacuolation in NASH. **(d)** The general course map visualizes general toxic courses common to multiple specific toxic courses (differentiated by colour).

3. Route search

In the development of pharmaceuticals, the requirements for a drug safety assessment system will differ depending on the development stage. In the early stages of drug discovery, developers want to know how the process will progress and whether a serious disorder will occur in the body. In contrast, in the pre-clinical stage, pathologists prefer to obtain information about the possible causes of an effect before observing the pathological findings.

The TOXPLOT system can provide both forward and backward route searches from a node of interest in a toxic course map. Figure 7c shows an example that compares the differences in retrospective analyses of lipid vacuolation as a pathological finding for phospholipidosis and NASH. In the course of NASH, vacuolation involves lipid accumulation, and the upstream process is dysfunction of fatty acid degradation and the like (Fig. 7c, upper part). In contrast, in the course of phospholipidosis, the vacuolation involves phospholipid accumulation due to phospholipid metabolism imbalance caused by dysfunction of phospholipid degradation, upstream of which is negative regulation of phospholipase-mediated phospholipid degradation caused by CAD (Fig. 7c lower part). Next, when comparing the route of downstream processes in each toxic course, the findings for NASH reveal that increasing hepatocellular volume involves ballooning, whereas in phospholipidosis, myelin figure in lysosome can be shown in each toxic course. Therefore, users can capture the differences in the specific pathological findings of each course. Hence, our map not only helps users search for the route of each toxic course, but it also helps them understand the differences between toxic courses given one common process or finding.

4. General course map

One problem in drug safety research is how to reduce unexpected toxicity. Therefore, it is important to provide comprehensive information about the causal relationship of mechanisms across multiple toxic courses. The TXPO provides a general course map that visualizes general toxic courses common to multiple specific toxic courses based on ontology (Fig. 7d). In drug safety evaluation, toxicologists sometimes want to know whether one phenomenon occurring in a particular toxic course could occur in other toxic courses. For instance, in the course of phospholipidosis, ‘phospholipid accumulation in lysosome’ can cause ‘increasing hepatocyte volume’. The TXPO system extracts information from the Resource Description Framework (RDF) database using SPARQL and automatically generates a general course map (Supplementary Information 4). In this map, common processes are represented as large nodes. Users can see that ‘increasing hepatocyte volume’ is common to other toxic courses such as cholestasis. Moreover, users can obtain information regarding different causes associated with other courses. Here, it is clear that ‘bile acid accumulation’ occurs specifically in the course of cholestasis and in other courses such as ‘smooth endoplasmic reticulum proliferation’ in ground glass appearance’.

Furthermore, by generalizing causes using the ontological hierarchy tree, the possible causes of increasing cell volume such as increases in the number of organelles or accumulation of intracellular substances can be classified. Cells have very specialized functions under normal conditions; however, in a toxic environment, they might increase their protective function through alternative pathways or perform alternative functions. By analysing both general and specific causal relationships, theoretically possible mechanisms for new paths could be found that could not be obtained by each course theme, which would improve drug risk discovery for drug safety evaluation.

Discussion

Concerning toxicity, there are many databases²⁸ such as DrugMatrix (<https://ntp.niehs.nih.gov/drugmatrix/index.html>) and the Liver Toxicity Knowledge Base (LTKB)²⁹, however, they focus on genomes or compounds. By employing systemic functional decomposition, the TXPO covers various processes across granularities in a consistent manner. We confirmed that we can describe both pathway- and molecular-level processes in a unified manner regarding ER stress. However, we found that the number of molecular processes is so large, it can be difficult to grasp the overall picture of the mechanism. Therefore, the TXPO deals primarily with process–process interactions with granularities starting from the organelle level. At the molecular level, we describe molecules as participants in toxic course processes. Furthermore, we explain the role of each molecule in a given specific process. Regarding in vitro screening, TOXPLOT can visualize marker genes and explicate the molecule roles in a toxic course, which will contribute to early risk management during drug discovery.

Understanding toxicity mechanisms is challenging. Among the many issues involved, one aspect is the complexity of various interactions in a toxic course. We demonstrated that our imbalance model can clarify the context and distinguish toxic actions from body defence functions at each granularity, thus facilitating the interpretations of toxic mechanisms.

In general, expert knowledge is hard for other experts to understand; for example, molecular biologists find it extremely difficult to distinguish findings in pathology, and vice versa, which fragments knowledge and makes it difficult to capture an overall picture of toxicities. The Comparative Toxicogenomics Database (CTD)³⁰ can be used to provide pathways and processes such as KEGG and GO. However, because each piece of information is provided independently, users cannot perceive the causal relationships of toxic processes. TXPO can provide systematized flexible information about molecules, pathways, and toxic processes through generalization, specialization, and other relationships in a consistent manner. Based on a philosophical view, ontology makes the intrinsic nature explicit. Using the systematized knowledge infrastructure of TXPO, TOXPLOT supports the interpretation of toxicity mechanisms, accelerates knowledge sharing, and realizes knowledge interoperability.

TXPO and TOXPLOT also help bridge the gap between basic science and clinical medicine. For example, ICD11 (<https://icd.who.int/browse11/l-m/en>) and Disease Ontology¹⁶ do not include the term ‘phospholipidosis’.

However, our ontology reveals that the sphingomyelin disorder, which is a type of phospholipidosis in toxicology, and Niemann–Pick types A and B have the following commonalities in their related processes:

- Structural commonality: lysosomes
- Causal relationship commonality: dysfunction of sphingomyelin degradation causes sphingomyelin accumulation in lysosomes
- Finding commonality: macrophage aggregation and myelin figures as myelin-like layered structures under electron microscopy

Regardless of whether they are internal or external factors, our system can provide causal relationships owing to dysfunction in the body system, homeostasis imbalance, and its outcomes. Our approach is hence able to transfer knowledge between toxicology and clinical medicine through commonalities. Therefore, from the perspectives of both toxicity and disease, TOXPILOT can support both drug risk management in drug development and therapeutic management in medicine. The imbalance model could determine a functioning performance level, which could help select patient groups with potential idiosyncratic toxicities. Because it includes latent toxicity, TOXPILOT will be useful for supporting patient care as a therapeutic approach in precision medicine. In this way, TOXPILOT enables knowledge sharing and generates interdisciplinary knowledge cycles based on the TXPO, and this could help elucidate complex mechanisms of toxicity.

We are currently annotating more toxic courses and enhancing the level of sophistication of the terms in the TXPO. In future, we plan to cover toxic courses in other organs. Recently, the TXPO joined the Open Biological and Biomedical Ontology (OBO) Foundry for knowledge sharing among not only toxicologists but also other biomedical communities. Via the OBO Foundry site (<http://www.obofoundry.org>), TXPO is available from the NCBO BioPortal, Aber-OWL (<http://aber-owl.net/ontology/TXPO>) and Ontobee (<http://www.ontobee.org/browser/index.php?o=txpo>).

A limitation of our basic principle of focusing on toxic processes is that we do not have much information about drug efficacy. To increase the comprehensiveness of knowledge in drug development, it may necessary to collaborate with other ontologies such as the Drug Target Ontology³¹. We are now extending valuable knowledge from various other ontologies, such as the Disease Ontology and Human Phenotype Ontology³². Bridging the gap in toxicity knowledge from the domains of basic science to clinical medicine could help elucidate multiple mechanisms of toxicity.

Methods

TXPO development. Using textbooks^{1,20,33–35}, we researched drug-induced hepatotoxic mechanisms and obtained information about toxic courses and related processes, molecules (and their roles), and biological structures. Next, we searched for the latest information from toxic course-related articles using the PubMed search terms listed in Supplementary Table S1.

We used the ontology editing tool Protégé 5.2.0³⁶ to develop the TXPO in the Web Ontology Language (OWL) and HermiT reasoner³⁷, which is a Protégé plug-in.

Supplementary Schema S1 shows examples of the TXPO development process, which consisted of the following steps. (1) Each toxic course was defined, and related information was annotated using the Annotation Properties. (2) The processes constituting each toxic course were described using a ‘has part’ relation as an Object Property. (3) Each process was generalized using an *is-a* hierarchy to reveal processes common to multiple toxic courses, biological processes, and biomedical-independent processes. (4) Each process was decomposed into sub-processes (using the ‘has part’ relation). (5) The biological structure in which the process takes place was described (using ‘occurs in’). (6) Molecules, drugs, and their roles in the process were defined. (7) Causal relationships between process were defined using a ‘has result’ relation.

When generalizing the *is-a* tree structure, we reused existing ontologies. Domain-independent general entities were based on BFO, and biomedical entities were imported manually from existing ontologies from the NCBO BioPortal³⁸. These biomedical ontologies include UBERON⁸, Cell Ontology⁹, NCBI Taxonomy¹⁰, ChEBI¹¹, Gene Ontology¹², PATO¹³, INOH¹⁴, and the OGG¹⁵.

TOXPILOT development. TOXPILOT consists of an ontology library, an RDF database, and a web application system (Supplementary Fig. S1). The TXPO file is stored in the ontology library, and the file is converted into RDF format, which represents data using a triplet of Subject, Predicate, and Object, by Protégé. The RDF data are then stored in an RDF store using Apache Jena Fuseki (<https://jena.apache.org/documentation/fuseki2/>) and a SPARQL endpoint is constructed. For the web application system for TOXPILOT, the necessary information is dynamically acquired via SPARQL queries. Moreover, TOXPILOT generates graphs using D3.js³⁹, which is a JavaScript library. TOXPILOT is publicly available online (<https://toxipilot.nibiohn.go.jp>). New term requests and issue reporting can be made via its GitHub tracker (<https://github.com/txpo-ontology/TXPO/issues>).

Received: 16 January 2020; Accepted: 7 August 2020

Published online: 03 September 2020

References

- Klaassen, C. D., Casarett, L. J. & Doull, J. *Casarett and Doull's Toxicology: The Basic Science of Poisons* 8th edn. (McGraw-Hill, New York, 2013).
- Arp, R., Smith, B. & Spear, A. D. *Building Ontologies with Basic Formal Ontology* (Massachusetts Institute of Technology, Cambridge, 2015).
- Mizoguchi, R. *Ontology Engineering* (Ohmsha, Tokyo, 2005).
- Kanehisa, M., Furumichi, M., Tanabe, M., Sato, Y. & Morishima, K. KEGG: new perspectives on genomes, pathways, diseases and drugs. *Nucleic Acids Res.* **45**, D353–D361. <https://doi.org/10.1093/nar/gkw1092> (2017).
- Slenter, D. N. *et al.* WikiPathways: a multifaceted pathway database bridging metabolomics to other omics research. *Nucleic Acids Res.* **46**, D661–D667. <https://doi.org/10.1093/nar/gkx1064> (2018).
- Jassal, B. *et al.* The reactome pathway knowledgebase. *Nucleic Acids Res.* <https://doi.org/10.1093/nar/gkz1031> (2019).
- Ankley, G. T. & Edwards, S. W. The adverse outcome pathway: a multifaceted framework supporting 21st century toxicology. *Curr. Opin. Toxicol.* **9**, 1–7. <https://doi.org/10.1016/j.cotox.2018.03.004> (2018).
- Mungall, C. J., Torniai, C., Gkoutos, G. V., Lewis, S. E. & Haendel, M. A. Uberon, an integrative multi-species anatomy ontology. *Genome Biol.* **13**, R5. <https://doi.org/10.1186/gb-2012-13-1-r5> (2012).
- Diehl, A. D. *et al.* The cell ontology 2016: enhanced content, modularization, and ontology interoperability. *J. Biomed. Semant.* **7**, 44. <https://doi.org/10.1186/s13326-016-0088-7> (2016).
- Federhen, S. The NCBI taxonomy database. *Nucleic Acids Res.* **40**, D136–143. <https://doi.org/10.1093/nar/gkr1178> (2012).
- Hastings, J. *et al.* ChEBI in 2016: improved services and an expanding collection of metabolites. *Nucleic Acids Res.* **44**, D1214–1219. <https://doi.org/10.1093/nar/gkv1031> (2016).
- Gene Ontology Consortium. Gene ontology consortium: going forward. *Nucleic Acids Res.* **43**, D1049–1056. <https://doi.org/10.1093/nar/gku1179> (2015).
- Gkoutos, G. V. *et al.* Ontologies for the description of mouse phenotypes. *Comput. Funct. Genomics* **5**, 545–551. <https://doi.org/10.1002/cfg.430> (2004).
- Yamamoto, S. *et al.* INOH: ontology-based highly structured database of signal transduction pathways. *Database* **2011**, 0052. <https://doi.org/10.1093/database/bar052> (2011).
- He, Y., Liu, Y. & Zhao, B. OGG: a biological ontology for representing genes and genomes in specific organisms. In *5th International Conference on Biomedical Ontology (ICBO2014)*, 13–20, Houston, Texas, USA, (2014).
- Bello, S. M. *et al.* Disease ontology: improving and unifying disease annotations across species. *Dis. Model Mech.* <https://doi.org/10.1242/dmm.032839> (2018).
- Kitamura, Y., Kashiwase, M., Fuse, M. & Mizoguchi, R. Deployment of an ontological framework of functional design knowledge. *Adv. Eng. Inform.* **18**, 115–117 (2004).
- Mizoguchi, R., Kitamura, Y. & Borgo, S. A unifying definition for artifact and biological functions. *Appl. Ontol.* **11**, 129–154. <https://doi.org/10.3233/AO-160165> (2016).
- Shayman, J. A. & Abe, A. Drug induced phospholipidosis: an acquired lysosomal storage disorder. *Biochim. Biophys. Acta* **1831**, 602–611. <https://doi.org/10.1016/j.bbali.2012.08.013> (2013).
- Strayer, D. S., Rubin, E., Saffitz, J. E. & Schiller, A. L. *Rubin's Pathology: Clinicopathologic Foundations of Medicine* 7th edn. (Wolters Kluwer Health, Alphen aan den Rijn, 2015).
- Mizoguchi, R. *et al.* River flow model of diseases. In *2nd International Conference on Biomedical Ontology (ICBO2011)*, 26–30, Buffalo, NY, USA (2011).
- Kumar, V., Abbas, A. K., Aster, J. C. & Robbins, S. L. *Robbins Basic Pathology* 9th edn. (Elsevier, Amsterdam, 2013).
- Canals, D., Perry, D. M., Jenkins, R. W. & Hannun, Y. A. Drug targeting of sphingolipid metabolism: sphingomyelinases and ceramidases. *Br. J. Pharmacol.* **163**, 694–712. <https://doi.org/10.1111/j.1476-5381.2011.01279.x> (2011).
- Luo, L. *et al.* Desipramine ameliorates Cr(VI)-induced hepatocellular apoptosis via the inhibition of ceramide channel formation and mitochondrial PTP opening. *Cell Physiol. Biochem.* **34**, 2128–2136. <https://doi.org/10.1159/000369657> (2014).
- Yamagata, Y., Igarashi, Y., Nakatsu, N. & Yamada, H. TXPO: A toxic process ontology for better understanding of drug-induced liver injury. In *10th International Conference on Biomedical Ontology (ICBO2019)*, Buffalo, NY, USA (2019).
- Yamagata, Y. *et al.* Development of liver toxicity ontology for drug safety evaluation and its application. *Trans. Jpn. Soc. Artif. Intell.* **34**(D-181), 81–18 (2019).
- Sawada, H., Takami, K. & Asahi, S. A toxicogenomic approach to drug-induced phospholipidosis: analysis of its induction mechanism and establishment of a novel in vitro screening system. *Toxicol. Sci.* **83**, 282–292. <https://doi.org/10.1093/toxsci/kfh264> (2005).
- Yamagata, Y., Yamada, H. & Horii, I. Current status and future perspective of computational toxicology in drug safety assessment under ontological intellection. *J. Toxicol. Sci.* **44**, 721–735. <https://doi.org/10.2131/jts.44.721> (2019).
- Chen, M. *et al.* FDA-approved drug labeling for the study of drug-induced liver injury. *Drug Discov. Today* **16**, 697–703. <https://doi.org/10.1016/j.drudis.2011.05.007> (2011).
- Davis, A. P. *et al.* The comparative toxicogenomics database: update 2019. *Nucleic Acids Res.* **47**, D948–D954. <https://doi.org/10.1093/nar/gky868> (2019).
- Lin, Y. *et al.* Drug target ontology to classify and integrate drug discovery data. *J. Biomed. Semant.* **8**, 50. <https://doi.org/10.1186/s13326-017-0161-x> (2017).
- Kohler, S. *et al.* Expansion of the Human Phenotype Ontology (HPO) knowledge base and resources. *Nucleic Acids Res.* **47**, D1018–D1027. <https://doi.org/10.1093/nar/gky1105> (2019).
- Nihon Tokishikoroji Gakkai. *Tokishikoroji* (Asakura Shoten, Hokota, 2009).
- Gakkai, N. D. B. *Shin dokusei byori soshikigaku* (Nishimurashotentokyohenshubu, Tokyo, 2017).
- Kaplowitz, N. & DeLeve, L. D. *Drug-Induced Liver Disease* (Academic Press, Cambridge, 2013).
- Musen, M. A. The Protege Project: a look back and a look forward. *AI Matters* **1**, 4–12. <https://doi.org/10.1145/2757001.2757003> (2015).
- Glimm, B., Horrocks, I., Motik, B., Stoilos, G. & Wang, Z. HermiT: an OWL 2 reasoner. *J. Autom. Reason.* **53**, 245–269. <https://doi.org/10.1007/s10817-014-9305-1> (2014).
- Whetzel, P. L. *et al.* BioPortal: enhanced functionality via new Web services from the National Center for Biomedical Ontology to access and use ontologies in software applications. *Nucleic Acids Res.* **39**, W541–545. <https://doi.org/10.1093/nar/gkr469> (2011).
- Bostock, M., Ogievetsky, V. & Heer, J. D3: data-driven documents. *IEEE Trans. Vis. Comput. Graph.* **17**, 2301–2309. <https://doi.org/10.1109/TVCG.2011.185> (2011).

Acknowledgements

This research is supported by the Japan Agency for Medical Research and Development (AMED) under Grant Number 19nk0101103h0005. The authors would like to express their thanks and appreciation to Dr Yoshinobu Igarashi, Dr Noriyuki Nakatsu (NIBIOHN), Dr Katsuhisa Horimoto, Dr Kazuhiko Fukui (AIST), and Professor Yoshihiro Uesawa (Meiji Pharmaceutical University), who took the time to discuss this research. The authors also would like to thank Nobutake Kato.

Author contributions

Y.Y. designed the study, developed the ontology, annotated the data, developed system, and wrote the manuscript. H.Y. started this project and contributed to data interpretation from a toxicological viewpoint. All authors reviewed and approved this version for publication.

Competing interests

The authors declare no competing interests.

Additional information

Supplementary information is available for this paper at <https://doi.org/10.1038/s41598-020-71370-7>.

Correspondence and requests for materials should be addressed to Y.Y. or H.Y.

Reprints and permissions information is available at www.nature.com/reprints.

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2020