

Research Article

Human Environmental Disease Network: A Computational Model to Assess Toxicology of Contaminants

Olivier Taboureau^{1,2,3} and Karine Audouze^{1,2}

¹INSERM UMR-S973, Molécules Thérapeutiques in silico, Paris, France; ²University of Paris Diderot, Paris, France;

³Novo Nordisk Foundation Center for Protein Research, Faculty of Health and Medical Sciences, University of Copenhagen, Copenhagen, Denmark

Summary

During the past decades, many epidemiological, toxicological and biological studies have been performed to assess the role of environmental chemicals as potential toxicants associated with diverse human disorders. However, the relationships between diseases based on chemical exposure rarely have been studied by computational biology.

We developed a human environmental disease network (EDN) to explore and suggest novel disease-disease and chemical-disease relationships. The presented scored EDN model is built upon the integration of systems biology and chemical toxicology using information on chemical contaminants and their disease relationships reported in the TDDb database. The resulting human EDN takes into consideration the level of evidence of the toxicant-disease relationships, allowing inclusion of some degrees of significance in the disease-disease associations. Such a network can be used to identify uncharacterized connections between diseases. Examples are discussed for type 2 diabetes (T2D). Additionally, this computational model allows confirmation of already known links between chemicals and diseases (e.g., between bisphenol A and behavioral disorders) and also reveals unexpected associations between chemicals and diseases (e.g., between chlordane and olfactory alteration), thus predicting which chemicals may be risk factors to human health.

The proposed human EDN model allows exploration of common biological mechanisms of diseases associated with chemical exposure, helping us to gain insight into disease etiology and comorbidity. This computational approach is an alternative to animal testing supporting the 3R concept.

Keywords: computational method, environmental contaminants, predictive toxicology, systems biology, human disease network

1 Introduction

Although it is well established that genes and environmental factors influence common human diseases, understanding disease-causing defects is still a challenge (Hunter, 2005). Comorbidity, i.e., the emergence of one or more additional disorders co-occurring with a primary disease due to a given treatment, lifestyle or environmental exposure, increases the degree of complexity.

Identifying how diseases are connected to each other would improve our understanding of the etiology and pathogenesis of similar diseases. Several network based approaches have allowed deciphering disease comorbidity (Goh et al., 2007; Lee et al., 2008; Hidalgo et al., 2009; Suthram et al., 2010;

Davis and Chawla, 2011; Sun et al., 2014a; Menche et al., 2015). The human “diseasome” was the first network-based model connecting diseases (Goh et al., 2007). The network was designed to associate diseases if they shared at least one gene, based on the Online Mendelian Inheritance in Man database (OMIM) (Hamosh et al., 2005). Recently, Roque et al. (2011) created an approach to gather phenotypic descriptions of patients from medical records that would suggest new disease-disease associations. Although all these studies provide comprehensive views of links between diseases, they all rely on existing knowledge, i.e., genes, pathways and phenotypic associations (Hidalgo et al., 2009).

Epidemiological and biological studies have suggested that a number of environmental chemicals may play a causative

Received July 20, 2016;
Accepted October 18, 2016;
Epub October 21, 2016;
<https://doi.org/10.14573/altex.1607201>



This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 International license (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution and reproduction in any medium, provided the original work is appropriately cited.



role in some human disorders. However, humans are potentially exposed to more than 80,000 substances for which little toxicity information exists. During recent years, toxicological and chemical databases have expanded substantially, and computational methods have been fine-tuned, so that *in silico* approaches to toxicity assessment now appear feasible and highly suitable (Knudsen et al., 2013; Kongsbak et al., 2014). The U.S. EPA (Knudsen et al., 2013) and a National Research Council expert committee (NRC, 2007) has recommended that *in silico* approaches should be included in future assessments of toxicity with the aim to reduce and refine existing methods. While *in silico* computer simulations cannot substitute biological testing, they can help focus on particular substances and targets to allow priority setting for more efficient testing. Overall, computational approaches can reduce animal experiments according to the 3R definition (Replacement, Reduction, Refinement).

Computational systems biology studies have shown up links between chemicals and diseases such as between the pesticide dichlorodiphenyltrichloroethane (DDT) and type II diabetes (Audouze and Grandjean, 2011), and between persistent organic pollutants (POPs) and metabolic diseases (Ruiz et al., 2016). However, the area has not been systematically screened, and environmental factors are rarely considered when creating disease-disease networks. Investigations reported in this area to date include a chemical-disease inference system based on chemical-protein interactions, ChemDIS, which was created with the aim to predict potential health risks associated with chemicals (Tung, 2015). Further, environmental etiological factors and genetic factor associations have been described using Medical Subject Headings (MeSH) terms (Liu et al., 2009). Also, a recent study has developed an integrated disease network based on various heterogeneous data types including disease-chemical associations (Sun et al., 2014b). Although interesting, such models are usually limited to binary data (chemical linked or not linked to a disease) and do not take into account the degree of the associations. For example, the severity of the chemical toxicity is rarely considered in a network-based approach and integration of such information would be valuable in the interpretation of the model outcomes.

In the present study, our main objective was to create a human environmental disease network (EDN) based on chemical-disease information from a comprehensive resource of chemicals and diseases, the Collaborative on Health and the Environment Toxicant and Disease Database (TDDB)¹. This database compiles existing information between chemical contaminants and approximately 200 human diseases based on biological and epidemiological evidence. Interestingly, it includes a level of evidence (from limited to good) between chemicals and diseases to estimate how the chemical exposure could contribute to the diseases. For example, mercury

is associated with cognitive impairment and behavioral problems with a high level of evidence, whereas methoxychlor is associated with reduced male fertility with limited evidence. To develop our EDN, we took advantage of the computational network biology approach that we developed previously with toxicogenomics data (Audouze et al., 2010). The concept, originally based on Protein-Protein Associations Network (P-PAN), can be transposed to other areas of application and has shown success with experimental validation of novel associations between chemicals and biological targets (Audouze et al., 2014). Although the list of chemicals and disease associations in the TDDB database is far from complete, it has the advantage of having three evidence layers. Therefore, instead of limiting ourselves to “binary” information (i.e., presence versus absence of chemical-disease associations), we have the opportunity to include some “weight” in the association and so to go beyond the bipartite network-based approach (Lee et al., 2008). Here we also present a case study to demonstrate the ability of our EDN model to predict disease-disease connections, chemical-disease links, and potential new associations. To our knowledge, this is the first time that a disease comorbidity network based on environmental chemicals has been developed using several layers of evidence.

2 Materials and methods

Data set

We extracted chemical-disease associations from the publicly available TDDB database (as of April 2015). The database contains information on 2790 connections between 601 environmental chemicals and 198 human diseases. We used the three available strengths of evidence to create a three-level disease network: The “strong evidence” (SE) represents a verified link between a chemical and a disease. Three cases fall under this category: (a) the chemical toxicity is well known and the chemical is recognized to cause the disease, (b) the causal associations have been found in recent, large, prospective or retrospective cohort studies, and (c) the chemicals are listed as group 1 human carcinogens by the International Agency for Research on Cancer (IARC)².

The “good evidence” (GE) includes chemical-disease associations based on epidemiological studies, chemicals listed in the IARC group 2A² (limited evidence for humans and strong for animal – probably carcinogenic) and chemicals listed by OEHHHA’s Prop 65 program³.

Finally, the “limited evidence” (LE) category contains chemicals associated with diseases based on case reports, on conflicting evidence, and chemicals listed in the IARC group 2B² and the EPA group B2⁴ (limited evidence in humans and in animals – possibly carcinogenic). Although the extracted data may be limited, the TDDB database represents the most

¹ <http://www.healthandenvironment.org/tddb>

² <http://monographs.iarc.fr/ENG/Classification/>

³ <https://oehha.ca.gov/>

⁴ <http://www.epa.gov/ttnatw01/toxsource/carcinogens.html>

complete repository of chemicals associated with human diseases with evidence information.

Generating a high confidence human disease-disease network

The relevant chemical-disease links collected from the TDDDB database were used to generate the disease-disease network. Taking all the information into account, the maximum number of diseases associated with a group of chemicals, i.e., pesticides, is 99, and the maximum number of chemicals associated with one disease, i.e., fetotoxicity, is 70. Looking only at the SE layer, the highest number of diseases (17 diseases) is linked to lead, and the highest number of chemicals (17 chemicals) is associated with the disease hepatitis. The EDN was created by representing each disease as a node, and linking any disease-disease pair for which at least one overlapping chemical was identified by an edge. The disease-disease pairs were converted into a non-redundant list of associations to develop our model, i.e., if diseases A and B are linked, the network may have two associations A-B and B-A. In our approach, only one of these pairs was retained to create the EDN.

Probabilistic score

To reduce noise and select the most significant disease-disease associations for prediction, we assigned a probabilistic score to each generated disease-disease pair. This score is based on the probability that a chemical linked to a disease A will also affect the other disease B. The diseases are represented by D_1, D_2, \dots, D_n .

For each disease, a set of chemicals is associated:

$$\text{Chem}_x := \{c \in \text{Chemicals} | c \text{ is associated with } D_x\}$$

This probabilistic score (pS) between a pair of diseases is calculated by the following equation:

$$\text{pS}(D_a, D_b) = \frac{|\text{Chem}_a \cap \text{Chem}_b|}{|\text{Chem}_a \cup \text{Chem}_b|}$$

The pS score increases with the strength of the association.

Exploration of the biological mechanisms: biological enrichment

To identify biological outcomes potentially related to individual chemicals, we first extracted known interactions between genes/proteins and chemicals using existing resources of information such as the ChemProt database (Kjærulff et al., 2013). Then, diseases and gene ontology (GO) information were integrated from two different sources in each gene/protein list. Gene-disease associations were extracted from the GeneCards database (March 2015), a comprehensive resource providing information on human genes and selected gene-related knowledge, such as functional and disease information (Safran et al., 2010). To investigate the GO information, all three GO categories, i.e., (a) molecular function, (b) biological processes, and (c) cellular components, were taken into consideration (Gene Ontology Consortium, 2015).

Diseases and GO terms enrichment analysis were finally performed with the gene's list for each analyzed chemical using a statistical test based on a hypergeometric distribution. A

significance level of 0.05 after Bonferroni correction for multiple testing was used to select the most relevant associations.

Predicting novel chemical-disease links

To predict diseases potentially linked to a chemical, a neighbor disease approach was performed based on the neighbor protein procedure described previously (Audouze et al., 2010). This approach is a multi-step procedure. First, diseases associated with the chemicals of interest from the TDDDB database are listed. These input diseases allow identifying network(s) surrounding them by using the network-neighbor's pull down approach (de Lichtenberg et al., 2005). In this procedure, the SE-EDN was queried for the input diseases, and associations between these were added. Next, the first order interactors of all input diseases were queried and added. A score was calculated for each neighbor, taking into account the topology of the surrounding network based on the ratio between total associations and associations with input diseases. Diseases with a score higher than the threshold (0.1), as defined previously, were kept in the final sub-network(s). This node inclusion parameter is at the conservative end of the optimal range for disease-disease interaction networks. Finally, within the aim to select all diseases' neighbors, all diseases in the network(s) were checked for associations among them, and the missing ones were added. A confidence score was established by testing each network for enrichment on the input set by comparing them against 1.0×10^{-4} random networks. The individual disease score was used to rank them, allowing prediction of diseases potentially linked to the chemical.

3 Results

3.1 Generating an environmental disease network (EDN)

Based on chemical-disease associations extracted from the TDDDB database, we constructed a human EDN using the three levels of evidence, i.e., SE, GE and LE. An overview of the strategy is shown in Figure 1. In total, the resulting EDN consists of 6258 associations between 196 diseases. The SE level contains 125 interconnected diseases, the GE level 141 diseases and the LE level 138 diseases. To reduce noise and select the most significant associations, we assigned a probabilistic score (pS) with each disease-disease association represented by the weight of each link.

3.2 Mining the environmental disease network

To simplify interpretation, each disease was classified into 19 primary disorder classes following the classification scheme described in the human disease network (Goh et al., 2007) (see Tab. S1, <https://doi.org/10.14573/altex.1607201s>). The classification is based on the biological systems affected by the diseases. For example, 32 diseases belong to a "cancer" class, 25 diseases constitute a "respiratory" class and only one disease represents the metabolism class. Interaction of two diseases belonging to the same class is defined as intra-class. Interaction of

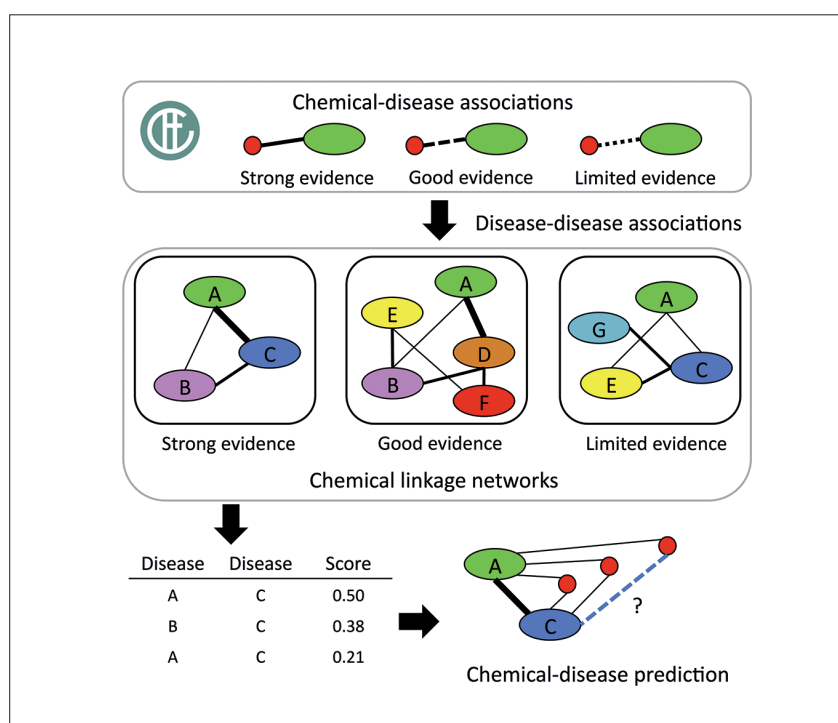


Fig. 1: Workflow of the proposed system's chemical biology strategy for predicting disease-disease and chemical-disease associations

Information on chemicals known to be linked to diseases and their evidence levels were extracted from the TDDb database and cleaned. The EDN model was then created using these data based on a protein-protein association network procedure, assuming that two nodes (i.e., diseases) are connected to each other if they share at least one chemical for which causal evidence is associated with both diseases. A probabilistic score was assigned to each disease pair in order to rank them. The higher the score, the stronger the association is. Using a network-neighbor's pull down procedure, prediction of connection between chemical and disease was performed.

two diseases from two different classes is defined as inter-class. When a disease affects several biological systems, only one class is assigned to the disease, based on the system known to be the most affected. Therefore, each disease is annotated to only one of the 19 classes.

Altogether, the EDN displays intra- and inter-class connections between 196 diseases. For example, Attention Deficit and Hyperactivity Disorder (ADHD) is connected to cognitive impairment, behavioral problems, low birth weight and fetotoxicity. Previous studies have shown association between prenatal mercury exposure and neuro-developmental disorders as behavioral diseases (Grandjean et al., 2014; Bellanger et al., 2015). This suggests that the environmental origins of diseases may be shared between diseases due to similar mechanisms of action of chemicals in complex disorders.

In a second step, we decided to explore the three levels of evidence on the EDN independently to evaluate the ability of the proposed approach to connect diseases according to these levels. In the SE layer, 125 diseases among the 196 appear connected via 1274 interactions. Figure 2 indicates that diseases in the SE layer tend to cluster by disease categories, except for cancer disorders, which are spread all over in the network. Not surprisingly, some well-known diseases are retrieved among the top intra-class associations, e.g., asthma-rhinitis; abnormal sperm-reduced male fertility; bronchitis chronic-chronic obstructive pulmonary disease (COPD); ADHD-cognitive impairment. The cancer-related disorders and the respiratory diseases appear to be the predominant classes including, respectively, 78% and 88% of the diseases involved in the EDN. The number of epidemiological studies performed on cancers and respiratory disorders may

explain the prominence of these diseases here. Two other classes are significantly represented in the SE, which are the developmental and the neurological classes (with 64% and 47% of the diseases). Among the top inter-class associations, we retrieved links between neurological diseases (cognitive impairment) and developmental disorders (low birth weight).

The two others layers, GE (Fig. S1, <https://doi.org/10.14573/altex.1607201s>) and LE (Fig. S2, <https://doi.org/10.14573/altex.1607201s>) show complementary information to the SE layer. The GE layer shows intra-class associations (coronary artery disease and hypertension, both being cardiovascular disorders), as well as inter-class links between reproductive and the developmental classes. In the GE layer, 2221 associations are displayed between 141 diseases. Among the most significant associations, ADHD is connected to color vision disturbance. Such links are supported by published studies, indicating that exposure to heavy metals may impair development of visual processing (Ethier et al., 2012).

The LE layer contains 3604 associations between 138 diseases. The most significant associations in this level are intra-class, and most of them concern cancer (breast cancer-lung cancer). For example, association is found between breast cancer and abnormal sperm, which is not surprising as both disorders have been suggested to be linked to several environmental chemicals (Reed et al., 2007). Overall, only few overlaps are retrieved, such as fetotoxicity and low birth weight, cognitive impairment and ADHD (significant associations).

To demonstrate the potential of these networks for the prediction of new disease-disease interactions, a case study with type II diabetes is described below.

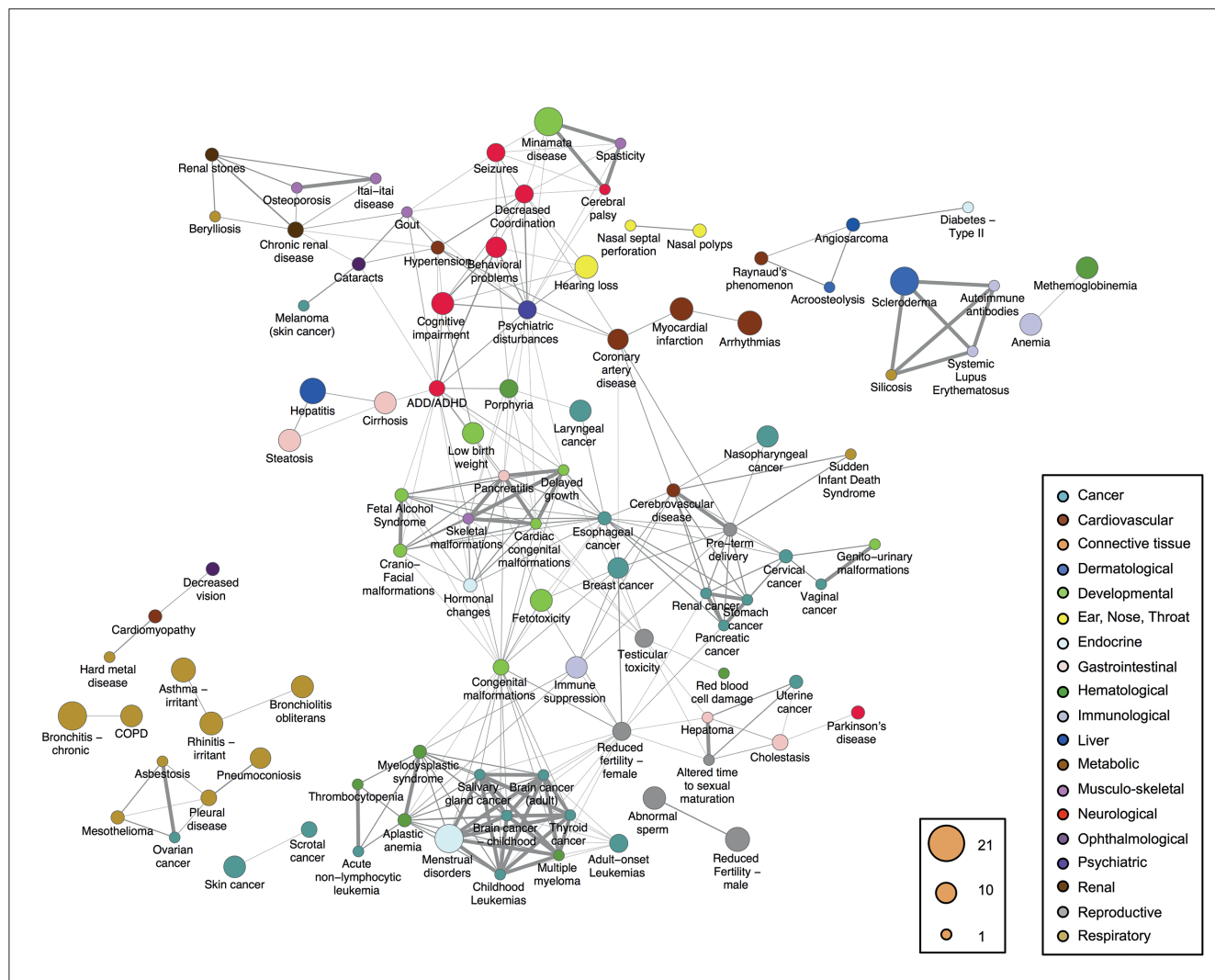


Fig. 2: Representation of the top significant disease-disease associations based on the probabilistic score using the strong level of evidence (SE)

Each node corresponds to a unique disease, colored according to the biological system (class) to which it belongs. The names of the 19 systems are shown on the right. Nodes' sizes are determined by the number of chemicals linked to the disease with a strong level of evidence. An edge is placed between two diseases if they share at least one chemical within the SE level. The width of an edge is proportional to the number of chemicals that are linked to both diseases. For example, six chemicals are linked with both myocardial infarction and arrhythmias disorders, resulting in an edge with a probabilistic score of 0.26. For reasons of clarity, only the top significant associations (based on the calculated score) are shown.

3.3 Case study: Type II diabetes

EDN exploration: prediction of T2D-disease associations

With the increasing number of diabetic people including children and young persons of reproductive age worldwide, there is a need to better understand potential secondary effects of the disease. To identify potential comorbidities between T2D and other diseases, we explored each level of the EDN independently. The three levels provided different information (Fig. 3).

In the LE layer, we can see associations between T2D and disorders of the reproductive system and hypertension. Associations between T2D and diseases of the reproductive

system, such as abnormal sperm and reduced male fertility, are linked via polychlorinated biphenyl compounds (PCBs). Studies have shown that T2D may affect male reproductive functions at multiple levels, including diminished sperm quality (Jangir and Jain, 2014). These hypotheses are also well in line with results provided by newly established computational tools. For example, HExpoChem, which allows prediction of diseases from chemical exposure (Taboureau et al., 2013) links the chemical PCB 126 to hypospadias via a protein complex of FGF9 ($p = 0.043$). On the ChemDis webserver (Tung, 2015), the association between PCB 126 and diabetes mellitus



SE	GE	LE	Disease name	Disease class
			Adult-onset leukemias	cancer
			Bladder cancer	cancer
			Brain cancer (adult)	cancer
			Breast cancer	cancer
			Colo-rectal cancer	cancer
			Gallbladder cancer	cancer
			Hepatocellular cancer	cancer
			Lung cancer	cancer
			Prostate cancer	cancer
			Skin cancer	cancer
			Stomach cancer	cancer
			Uterine cancer	cancer
			Arrhythmias	cardiovascular
			Coronary artery disease	cardiovascular
			Dyslipidemia	cardiovascular
			Hypertension	cardiovascular
			Contact dermatitis- Irritant	dermatological
			Hyperkeratosis	dermatological
			Skin ulceration	dermatological
			Cardiac congenital malformations	developmental
			Congenital malformations	developmental
			Genito-urinary malformations	developmental
			Oral clefts	developmental
			Hearing loss	Ear, nose, throat

SE	GE	LE	Disease name	Disease class
			Early onset menopause	endocrine
			Hormonal changes	endocrine
			Menstrual disorders	endocrine
			Thyroid disorders	endocrine
			Aplastic anemia	hematological
			Hodgkin's disease	hematological
			Multiple myeloma	hematological
			Porphyria	hematological
			Thrombocytopenia	hematological
			Angiosarcoma	liver
			Hepatitis	liver
			ADD/ADHD	neurological
			Alzheimer	neurological
			Parkinson	neurological
			Peripheral neuropathy	neurological
			Cataracts	ophthalmological
			Macular degeneration	ophthalmological
			Abnormal sperm	reproductive
			Altered sex ratio	reproductive
			Altered time to sexual maturation	reproductive
			Reduced fertility-female	reproductive
			Reduced fertility-male	reproductive

Fig. 3: Full prediction of T2D-disease associations within the three levels of evidence

To identify potential comorbidities between T2D and other diseases, each level of the EDN was independently explored, providing different information. Each biological system is depicted by a specific color indicating to which class a disease belongs.

is statistically inferred from a chemical-protein-disease relationship ($p = 1.98 \times 10^{-2}$).

Regarding the association of T2D and hypertension, published studies have supported a substantial overlap between T2D and hypertension in etiology and disease mechanisms (Gress et al., 2000; Cheung and Li, 2012). This interaction has been observed in diabetic patients who presented hypertension symptoms (Jensen et al., 2012). In our approach, both diseases are connected by exposure to PCBs. It has been reported independently that PCBs may increase insulin resistance, cause T2D (Kouznetsova et al., 2007) and are associated with hypertension (Everett et al., 2008). However, the underlying mechanism(s) remains to be ascertained. Using HExpoChem, PCB 126 was associated, though not statistically significantly, with insulin sensitivity via a protein complex of ADIPOQ ($p > 0.05$), and with arterial hypertension via a protein complex of AGTR1 and another complex of NOS3 (both $p > 0.05$). According to the Chemical Toxicogenomics Database (CTD)

(Grondin et al., 2016), these two diseases have 28 common genes and 42 common chemicals including PCBs (information from curated data), suggesting a potential similarity in the mechanism of action.

Fewer associations were identified in the two other layers (SE and GE). For example, a link between T2D and Hodgkin's disease (HD) is indicated via endocrine disruptor chemicals in GE (Fig. 3). This association is supported by a study that has examined epidemiological associations between T2D and the risk of HD, and concluded that the diseases may be linked (Mitri et al., 2008). The incidence of both T2D and HD has increased significantly during the past decades, and a study on a population-based cohort of more than 130 000 adults suggests that T2D may be associated with an increased risk of developing HD (Yang et al., 2016).

Hormonal changes and T2D are also predicted to be connected via two levels of the EDN (GE and LE). From a systems chemical biology perspective, T2D is linked to hormonal

Tab. 1: Biological enrichment for 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) and dichlorodiphenyltrichloroethane (DDT)
Diseases from the GeneCards databases and Gene Ontologies (GO) terms were considered. Bonferroni corrected p-values and the genes associated with diseases and GO terms are listed by HUGO gene symbol.

Chemical	Biological	Enrichment	No. of genes	p-value	Gene list
TCDD	Disease	Diabetes mellitus	26	1.90 ^{e-19}	CPT1A; EDN1; AKT2; GCK; HMOX1; HNF4A; HP; IRS1; KCNJ11; LEP; LEPR; NFKB1; ENPP1; PPARA; PPARG; RETN; PTPN1; SLC2A1; SLC2A2; SLC2A4; TNFRSF1A; C3; UCP2; WFS1; CAT; ADIPOQ
	GO process	regulation of hormone secretion (go:0046883)	14	3.36 ^{e-11}	CPT1A; EDN1; GCK; HNF4A; IL6; IRS1; KCNJ11; LEP; SLC2A1; SLC2A2; HNF1A; TCF7L2; IRS2; ADIPOQ
		response to steroid hormone stimulus (go:0048545)	9	0.032	EDN1; HMOX1; FAS; IL6; KCNJ11; PPARA; PPARG; TNFRSF1A; ADIPOQ
DDT	Disease	Diabetes mellitus	16	0.799	ADRB3; ESR2; ALB; G6PD; GGT1; GPT; IL1R1; NFKB1; PON1; MAPK8; RBP4; SHBG; SOD1; TNFRSF1A; CAT; PDE5A
	GO process	steroid hormone receptor activity (go:0003707)	10	1.59 ^{e-05}	NR0B1; ESR1; ESR2; HNF4G; AR; PGR; RXRB; NR0B2; NR1I2; NR1I3
		hormone activity (go:0005179)	4	n.s.*	FSHB; GNRH1; POMC; TSHB

* n.s., non-significant

changes via 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (GE layer) and dichlorodiphenyltrichloroethane (DDT) (LE layer). Recent computational studies have explored possible pathogenetic links between environmental chemicals and diseases through various data types, such as genome-wide associations and disease similarities, and found a potential link between both T2D and TCDD, and T2D and DDT (Audouze and Grandjean, 2011).

The EDN also shows that dermatological disorders may be linked to T2D (SE level), which is not surprising as skin complications related to T2D are common.

Exploration of the biological mechanisms: understanding the findings for T2D-disease predictions

To gain a better comprehension of the predicted associations between T2D and hormonal changes and decipher a possible biological relevance, we went one step further. We performed biological enrichment in order to suggest potential mode(s) of action of the two chemicals (TCDD and DDT) to understand their connections to both disorders. To identify biological outcomes, disease and GO enrichments were done on the protein lists extracted from the ChemProt database for TCDD and DDT.

In the disease database, 206 genes were connected to T2D. Among them, 26 are perturbed by TCDD, giving a significant adjusted $p = 1.904 \times 10^{-19}$. From the GO process, among the 14,650

genes, 162 were linked to regulation of hormone secretion. 14 of them are associated with TCDD, giving a $p = 3.361 \times 10^{-11}$. Another GO process, response to steroid hormone stimulus (294 genes among the 14,650), shows also a significant p-value of 0.032 via 9 genes linked to TCDD. So, we can see that several genes linked to TCDD are present in T2D and hormonal changes (Tab. 1).

Regarding potential links between DDT and T2D, 16 genes associated with DDT were known to be linked to diabetes mellitus in the disease database. After enrichment, DDT was significantly associated to steroid hormone receptor activity via 10 genes ($p = 1.596 \times 10^{-5}$) based on information from the GO function, which contains 15,209 genes (Tab. 1). Overall, these analyses support the findings linking T2D and hormonal changes identified by exploring the GE levels on the EDN.

3.4 Deciphering possible links between novel chemicals and disease

Besides revealing connections between diseases, the EDN can be used to assess the risk of a chemical to induce diseases. As chemicals may interact with several proteins or protein complexes (Paolini et al., 2006), and diseases may also be connected to multiple proteins, this approach can be helpful to identify potential relationships between chemicals and diseases. We therefore developed and applied a neighbor disease



Tab. 2: Mining the full EDN for diseases associated with bisphenol A (BPA), chlordane and hexachlorobenzene (HCB)

The number of diseases already known from the database to be associated with the chemical is shown, and the highest and lowest probabilistic scores are mentioned with the diseases. Predicted diseases are shown with their corresponding scores. For example, BPA is predicted to be connected to eight already known diseases, for which menstrual disorders has the highest score. Moreover, BPA is also predicted to be linked to five other diseases, among them leukemia, which has the highest score, and for which there is no literature support to date.

Environmental exposure	Known (all layers)	Diseases (max and min score)	Predicted diseases	Score
BPA	8	Menstrual disorders (1.65)	Leukemia childhood	2.57
		Altered time to sexual maturation (0.42)	Leukemia adult	2.09
			Breast cancer	1.72
			Testicular toxicity	1.65
			ADD/ADHD	1.30
Chlordane	11	Adult onset leukemia (2.95)	Brain cancer childhood	2.65
		Peripheral neuropathy (0.17)	Brain cancer adult	2.65
			Congenital malformation	2.33
			Fetotoxicity	1.89
			Olfactory alteration	0.48
HCB	15	Brain cancer adult (3.45)	Brain cancer child	4.49
		Hyperkeratosis (0.08)	Skin cancer	3.24
			Congenital malformation	3.23
			Abnormal sperm	1.77
			Lung cancer	1.52

procedure built on a neighbor protein procedure, which scores the associations between diseases, to the EDN model. The performance of this approach has been shown in previous studies (Audouze et al., 2010, 2014). Only the SE layer was used. Diseases known to be connected to various chemicals were listed independently, i.e., one disease list for each chemical. Each disease list was scanned into the disease network in order to identify other diseases potentially linked (with a high score) to the chemical (Tab. 2).

As a first example, we screened bisphenol A (BPA), which is an environmental estrogen used in the manufacture of polycarbonate plastics and epoxy resins to make food and beverage packaging. The use of BPA in food and beverage packaging has been banned in several countries due to its potential toxic effects on the reproductive system, and also on hypertension and metabolic disorders (Ariemma et al., 2016). The effect of BPA exposure on human brain and behavior is a relatively new issue, and particular concerns have been raised about its potential impact on children (Perez-Lobato et al., 2016). Using our approach, several diseases were identified as potentially linked to BPA (Tab. 2). For example, increasing evidence suggests

that BPA mimics estrogens in the body and might be associated with putative markers of breast cancer risk (McGuinn et al., 2015). Until now, no direct epidemiological associations exist, but it could well contribute as it affects the genes that defend against cancer (Bhan et al., 2014).

A relationship between BPA and behavioral disorders is also found in the network. Published studies suggest potential associations of prenatal/early life BPA exposure with behavior problems, including anxiety, depression, and ADHD in children (Mustieles et al., 2015; Casas et al., 2015). Also, boys exposed to higher BPA concentrations as a fetus or during early childhood were more likely to suffer from anxiety, aggression, depression and hyperactivity at age 7 (Harley et al., 2013).

We looked at chlordane as a second example. Chlordane is a pesticide banned in Europe and the US, but very persistent in the environment⁵. Therefore, exposure to chlordane is still harming the health of millions of people (Evangelou et al., 2016). The acute (short-term) effects of chlordane in humans consist of gastrointestinal distress and neurological symptoms, such as tremors and convulsions (National Library of Medicine, 1992). Chronic (long-term) inhalation exposure of

⁵ Stockholm Convention, UNEP, 2016: <http://chm.pops.int/TheConvention/ThePOPs/ListingofPOPs>

humans to chlordane results in effects on the nervous system (Kim et al., 2015). Our approach predicts chlordane to have an association with brain cancer, which is in line with the U.S. EPA classification as a Group B2 probable human carcinogen⁶. Chlordane was also predicted to be associated with olfactory alteration. Although a direct olfactory impairment from chemical exposure including pesticides has not been identified, it is known that some environmental chemicals may induce respiratory inflammations that cause such damage (Doty, 2015). Furthermore, it has been suggested that olfactory loss can occur as a result of exposure to chemicals present in air pollution or workplace situations (Quandt et al., 2016), but no specific link between chlordane and olfactory disorders has been reported.

We finally explored the synthetic industrial chemical hexachlorobenzene (HCB). HCB is a bioaccumulative, persistent and toxic pollutant defined by the Stockholm Convention⁵. Historically HCB was commonly used as a pesticide and fungicide⁷. Although HCB production has stopped in many countries⁷, the compound is still generated inadvertently, as a byproduct and/or impurity in the manufacture of various chlorinated compounds, and released into the environment (Mrema et al., 2012). In our network, HCB is predicted to be associated with reproductive disorders such as congenital malformation and abnormal sperm. Environmental exposure to endocrine disrupting chemicals, including HCB, have been suggested as a risk factor for male genital abnormalities such as hypospadias (Rignell-Hydbom et al., 2012; Krysiak-Baltyn et al., 2012). A recent study showed for the first time a correlation between serum concentration of HCB and semen quality (Paoli et al., 2015). Regarding its potential links to cancer, these results are supported by the International Agency for Research on Cancer and the U.S. EPA that classify HCB as a probable human carcinogen (U.S. EPA, 1999).

4 Discussion

The proposed approach offers a network-based hypothesis for the emergence of complex diseases, which cannot always be explained by genetic variability only, but also may be linked to environmental factor exposure.

Although our EDN can be of help in the understanding of disease co-occurrences, of mechanisms of action, and in the risk assessment of new chemicals, we are aware that the chemical-disease annotations used in this work are limited in terms of diseases and chemicals. For example, the version of the TDBB database used here does not include data on obesity or inflammatory bowel diseases. Similarly, the list of contaminants is relatively general for some classes of chemicals, e.g., air pollution and dusts. Therefore, we considered only a part of the currently available information in our EDN based-predic-

tion and an extension of the proposed network with integration of other data would be beneficial. Some efforts are on-going in measuring potential human exposure to environmental chemicals and facilitating public access to these data. As examples, we can mention the exposure data collection of the U.S. EPA⁶, the human exposome project⁸, the Heals project, which is the largest research project in Europe on environment and health⁹, and the national report on human exposure to environmental chemicals by the Centers for Disease Control and Prevention¹⁰.

Still, the TDBB database has the benefit of organizing the chemical-disease associations based on three levels of evidence that allowed us to develop an environmental disease network with a better selectivity than a global network. Developing biological networks with more selective data allows generating more accurate and predictive models. Another advantage of the proposed network-based approach is the ability to identify potential new chemical-disease relationships without taking into consideration the chemical structure as the majority of computational tools do in this area.

The next challenge will be to integrate further databases suitable for the generation of computational methods able to decipher potential risks associated with chemicals and to generate hypotheses, accelerating the hazard identification process. One way to screen more of the available data would be to use advanced text mining tools, such as one used to extract drug-adverse event information from electronic medical records (Roitmann et al., 2014). Crossing the hypotheses made by our approach with some other observations described in the literature could further improve the characterization of potential chemical-disease relationships. For example, a group led by Leonardo Trasande has recently developed a system to estimate health and economic costs related to endocrine disrupting chemicals (EDCs) exposure in the European Union (Trasande et al., 2015, 2016) and the USA (Attina et al., 2016). To estimate costs, they used available epidemiological and toxicological evidence for each EDC and weighted them. Such probability of causation, e.g., EDC causation of IQ loss and association with autism, childhood obesity or male infertility, could be crossed with other computational models and the TDBB database.

5 Conclusions

Despite all recent advances in high throughput interactome mapping and in disease gene identification, both the protein-protein interactions and our knowledge of disease associated genes remain incomplete (Menche et al., 2015). We present in this study a disease-disease network based on a degree of evidence from chemical-disease relationships. The ability of the EDN to identify novel disease-disease associations and chemical-dis-

⁶ <https://www.epa.gov>

⁷ Agency for Toxic Substances and Disease Registry - ATSDR report, 2015. <https://www.atsdr.cdc.gov>

⁸ <http://humanexposomeproject.com/>

⁹ <http://www.heals-eu.eu/>

¹⁰ <https://www.cdc.gov/exposurereport/>



ease associations was illustrated by several examples, such as the suggestion of links between BPA and ADHD, and chlordane and olfactory dysfunction, which would need further investigations to be confirmed. Such computational methods can be used to enhance our current knowledge and may help to prioritize further experimental testing.

References

- Ariemma, F., D'Esposito, V., Liguoro, D. et al. (2016). Low-dose bisphenol-A impairs adipogenesis and generates dysfunctional 3T3-L1 adipocytes. *PLoS One* 11, e0150762. <https://doi.org/10.1371/journal.pone.0150762>
- Attina, T. M., Hauser, R., Sathyanarayana S. et al. (2016). Exposure to endocrine-disrupting chemicals in the USA: A population-based disease burden and cost analysis. *Lancet Diabetes Endocrinol* 12, 996-1003. [https://doi.org/10.1016/S2213-8587\(16\)30275-3](https://doi.org/10.1016/S2213-8587(16)30275-3)
- Audouze, K., Juncker, A. S., Roque, F. J. S. S. A. et al. (2010). Deciphering diseases and biological targets for environmental chemicals using toxicogenomics networks. *PLoS Comput Biol* 6, e1000788. <https://doi.org/10.1371/journal.pcbi.1000788>
- Audouze, K. and Grandjean, P. (2011). Application of computational systems biology to explore environmental toxicity hazards. *Environ Health Perspect* 119, 1754-1759. <https://doi.org/10.1289/ehp.1103533>
- Audouze, K., Tromelin, A., Le Bon, A. M. et al. (2014). Identification of odorant-receptor interactions by global mapping of the human odorome. *PLoS One* 9, e93037. <https://doi.org/10.1371/journal.pone.0093037>
- Bellanger, M., Demeneix, B., Grandjean, P. et al. (2015). Neurobehavioral deficits, diseases, and associated costs of exposure to endocrine-disrupting chemicals in the European Union. *J Clin Endocrinol Metab* 100, 1256-1266. <https://doi.org/10.1210/jc.2014-4323>
- Bhan, A., Hussain, I., Ansari, K. I. et al. (2014). Bisphenol A and diethylstilbestrol exposure induces the expression of breast cancer associated long noncoding RNA HOTAIR in vitro and in vivo. *J Steroid Biochem Mol Biol* 141, 160-170. <https://doi.org/10.1016/j.jsbmb.2014.02.002>
- Casas, M., Forns, J., Martínez, D. et al. (2015). Exposure to bisphenol A during pregnancy and child neuropsychological development in the INMA-Sabadell cohort. *Environ Res* 142, 671-679. <https://doi.org/10.1016/j.envres.2015.07.024>
- Cheung, B. M. Y. and Li, C. (2012). Diabetes and hypertension: Is there a common metabolic pathway? *Curr Atheroscler Rep* 14, 160-166. <https://doi.org/10.1007/s11883-012-0227-2>
- Davis, D. A. and Chawla, N. V. (2011). Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. *PLoS One* 6, e22670. <https://doi.org/10.1371/journal.pone.0022670>
- de Lichtenberg, U., Jensen, L. J., Brunak, S. and Bork, P. (2005). Dynamic complex formation during the yeast cell cycle. *Science* 307, 724-727. <https://doi.org/10.1126/science.1105103>
- Doty, R. L. (2015). Neurotoxic exposure and impairment of the chemical senses of taste and smell. *Handb Clin Neurol* 131, 299-324. <https://doi.org/10.1016/B978-0-444-62627-1.00016-0>
- Ethier, A.-A., Muckle, G., Bastien, C. et al. (2012). Effects of environmental contaminant exposure on visual brain development: A prospective electrophysiological study in school-aged children. *Neurotoxicology* 33, 1075-1085. <https://doi.org/10.1016/j.neuro.2012.05.010>
- Evangelou, E., Ntritsos, G., Chondrogiorgi, M. et al. (2016). Exposure to pesticides and diabetes: A systematic review and meta-analysis. *Environ Int* 91, 60-68. <https://doi.org/10.1016/j.envint.2016.02.013>
- Everett, C. J., Mainous, A. G., Frithsen, I. L. et al. (2008). Association of polychlorinated biphenyls with hypertension in the 1999-2002 national health and nutrition examination survey. *Environ Res* 108, 94-97. <https://doi.org/10.1016/j.envres.2008.05.006>
- Gene Ontology Consortium (2015). Gene ontology consortium: Going forward. *Nucleic Acids Res* 43, D1049-1056. <https://doi.org/10.1093/nar/gku1179>
- Goh, K.-I., Cusick, M. E., Valle, D. et al. (2007). The human disease network. *Proc Natl Acad Sci* 104, 8685-8690. <https://doi.org/10.1073/pnas.0701361104>
- Grandjean, P., Weihe, P., Debes, F. et al. (2014). Neurotoxicity from prenatal and postnatal exposure to methylmercury. *Neurotoxicol Teratol* 43, 39-44. <https://doi.org/10.1016/j.ntt.2014.03.004>
- Gress, T. W., Nieto, F. J., Shahar, E. et al. (2000). Hypertension and antihypertensive therapy as risk factors for type 2 diabetes mellitus. Atherosclerosis risk in communities study. *N Engl J Med* 342, 905-912. <https://doi.org/10.1056/NEJM200003303421301>
- Grondin, C. J., Davis, A. P., Wieggers, T. C. et al. (2016). Advancing exposure science through chemical data curation and integration in the comparative toxicogenomics database. *Environ Health Perspect* 124, 1592-1599. <https://doi.org/10.1289/EHP174>
- Hamosh, A., Scott, A. F., Amberger, J. S. et al. (2005). Online Mendelian inheritance in man (OMIM), a knowledgebase of human genes and genetic disorders. *Nucleic Acids Res* 33, D514-517. <https://doi.org/10.1093/nar/gki033>
- Harley, K. G., Gunier, R. B., Kogut, K. et al. (2013). Prenatal and early childhood bisphenol A concentrations and behavior in school-aged children. *Environ Res* 126, 43-50. <https://doi.org/10.1016/j.envres.2013.06.004>
- Hidalgo, C. A., Blumm, N., Barabási, A.-L. and Christakis, N. A. (2009). A dynamic network approach for the study of human phenotypes. *PLoS Comput Biol* 5, e1000353. <https://doi.org/10.1371/journal.pcbi.1000353>
- Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nat Rev Genet* 6, 287-298. <https://doi.org/10.1038/nrg1578>

- Jangir, R. N. and Jain, G. C. (2014). Diabetes mellitus induced impairment of male reproductive functions: A review. *Curr Diabetes Rev* 10, 147-157. <https://doi.org/10.2174/1573399810666140606111745>
- Jensen, P. B., Jensen, L. J. and Brunak, S. (2012). Mining electronic health records: Towards better research applications and clinical care. *Nat Rev Genet* 13, 395-405. <https://doi.org/10.1038/nrg3208>
- Kim, S.-A., Lee, Y. M., Lee, H. W. et al. (2015). Greater cognitive decline with aging among elders with high serum concentrations of organochlorine pesticides. *PLoS One* 10, 1-9. <https://doi.org/10.1371/journal.pone.0130623>
- Kjærulff, S. K., Wich, L., Kringelum, J. et al. (2013). ChemProt-2.0: Visual navigation in a disease chemical biology database. *Nucleic Acids Res* 41, D464-469. <https://doi.org/10.1093/nar/gks1166>
- Knudsen, T., Martin, M., Chandler, K. et al. (2013). Predictive models and computational toxicology. *Methods Mol Biol* 947, 343-374. https://doi.org/10.1007/978-1-62703-131-8_26
- Kongsbak, K., Vinggaard, A. M., Hadrup, N. and Audouze, K. (2014). A computational approach to mechanistic and predictive toxicology of pesticides. *ALTEX* 31, 11-22. <https://doi.org/10.14573/altex.1304241>
- Kouznetsova, M., Huang, X., Ma, J. et al. (2007). Increased rate of hospitalization for diabetes and residential proximity of hazardous waste sites. *Environ Health Perspect* 115, 75-79. <https://doi.org/10.1289/ehp.9223>
- Krysiak-Baltyn, K., Toppari, J., Skakkebaek, N. E. et al. (2012). Association between chemical pattern in breast milk and congenital cryptorchidism: Modelling of complex human exposures. *Int J Androl* 35, 294-302. <https://doi.org/10.1111/j.1365-2605.2012.01268.x>
- Lee, D.-S., Park, J., Kay, K. A. et al. (2008). The implications of human metabolic network topology for disease comorbidity. *Proc Natl Acad Sci* 105, 9880-9885. <https://doi.org/10.1073/pnas.0802208105>
- Liu, Y. I., Wise, P. H. and Butte, A. J. (2009). The “etiome”: Identification and clustering of human disease etiological factors. *BMC Bioinformatics* 10, Suppl 2, S14. <https://doi.org/10.1186/1471-2105-10-S2-S14>
- McGuinn, L. A., Ghazarian, A. A., Joseph Su, L. and Ellison, G. L. (2015). Urinary bisphenol A and age at menarche among adolescent girls: Evidence from NHANES 2003-2010. *Environ Res* 136, 381-386. <https://doi.org/10.1016/j.envres.2014.10.037>
- Menche, J., Sharma, A., Kitsak, M. et al. (2015). Disease networks. Uncovering disease-disease relationships through the incomplete interactome. *Science* 347, 1257601. <https://doi.org/10.1126/science.1257601>
- Mitri, J., Castillo, J. and Pittas, A. G. (2008). Diabetes and risk of Non-Hodgkin's lymphoma: A meta-analysis of observational studies. *Diabetes Care* 31, 2391-2397. <https://doi.org/10.2337/dc08-1034>
- Mrema, E. J., Rubino, F. M., Brambilla, A. et al. (2012). Persistent organochlorinated pesticides and mechanisms of their toxicity. *Toxicology* 307, 74-88. <https://doi.org/10.1016/j.tox.2012.11.015>
- Mustieles, V., Pérez-Lobato, R., Olea, N., Fernández, M. F. (2015). Bisphenol A: Human exposure and neurobehavior. *Neurotoxicology* 49, 174-184. <https://doi.org/10.1016/j.neuro.2015.06.002>
- National Library of Medicine (1992). Hazardous Substances Databank. TOXNET, Medlars Management Section, Bethesda, MD, USA. <https://toxnet.nlm.nih.gov/cgi-bin/sis/htmlgen?HSDB>
- NRC – National Research Council (2007). *Toxicity Testing in the 21st Century: A Vision and a Strategy*. Washington, DC: National Academy Press.
- Paoli, D., Giannandrea, F., Gallo, M. et al. (2015). Exposure to polychlorinated biphenyls and hexachlorobenzene, semen quality and testicular cancer risk. *J Endocrinol Invest* 38, 745-752. <https://doi.org/10.1007/s40618-015-0251-5>
- Paolini, G. V., Shapland, R. H. B., van Hoorn, W. P. et al. (2006). Global mapping of pharmacological space. *Nat Biotechnol* 24, 805-815. <https://doi.org/10.1038/nbt1228>
- Perez-Lobato, R., Mustieles, V., Calvente, I. et al. (2016). Exposure to bisphenol A and behavior in school-age children. *Neurotoxicology* 53, 12-19. <https://doi.org/10.1016/j.neuro.2015.12.001>
- Quandt, S. A., Walker, F. O., Talton, J. W. et al. (2016). Olfactory function in Latino farmworkers: Subclinical neurological effects of pesticide exposure in a vulnerable population. *J Occup Environ Med* 58, 248-253. <https://doi.org/10.1097/JOM.0000000000000672>
- Rignell-Hydbom, A., Lindh, C. H., Dillner, J. et al. (2012). A nested case-control study of intrauterine exposure to persistent organochlorine pollutants and the risk of hypospadias. *PLoS One* 7, e44767. <https://doi.org/10.1371/journal.pone.0044767>
- Roitmann, E., Eriksson, R. and Brunak, S. (2014). Patient stratification and identification of adverse event correlations in the space of 1190 drug related adverse events. *Front Physiol* 5, 332. <https://doi.org/10.3389/fphys.2014.00332>
- Roque, F. S., Jensen, P. B., Schmock, H. et al. (2011). Using electronic patient records to discover disease correlations and stratify patient cohorts. *PLoS Comput Biol* 7, e1002141. <https://doi.org/10.1371/journal.pcbi.1002141>
- Ruiz, P., Perlina, A., Mumtaz, M. and Fowler, B. A. (2016). A systems biology approach reveals converging molecular mechanisms that link different POPs to common metabolic diseases. *Environ Health Perspect* 124, 1034-1041. <https://doi.org/10.1289/ehp.1510308>
- Safran, M., Dalah, I., Alexander, J. et al. (2010). GeneCards Version 3: The human gene integrator. *Database J Biol Databases Curation* 2010, baq020. <https://doi.org/10.1093/database/baq020>
- Sun, K., Buchan, N., Larminie, C. and Pržulj, N. (2014a). The integrated disease network. *Integr Biol (Camb)* 6, 1069-1079. <https://doi.org/10.1039/c4ib00122b>



- Sun, K., Gonçalves, J. P., Larminie, C. and Przulj, N. (2014b). Predicting disease associations via biological network analysis. *BMC Bioinformatics* 15, 304. <https://doi.org/10.1186/1471-2105-15-304>.
- Suthram, S., Dudley, J. T., Chiang, A. P. et al. (2010). Network-based elucidation of human disease similarities reveals common functional modules enriched for pluripotent drug targets. *PLoS Comput Biol* 6, e1000662. <https://doi.org/10.1371/journal.pcbi.1000662>
- Taboureau, O., Jacobsen, U. P., Kallhauge, C. et al. (2013). HExpoChem: A systems biology resource to explore human exposure to chemicals. *Bioinformatics* 29, 1231-1232. <https://doi.org/10.1093/bioinformatics/btt112>
- Trasande, L., Zoeller, R. T., Hass, U. et al. (2015). Estimating burden and disease costs of exposure to endocrine-disrupting chemicals in the European Union. *J Clin Endocrinol Metab* 100, 1245-1255. <https://doi.org/10.1210/jc.2014-4324>
- Trasande, L., Zoeller, R. T., Hass, U. et al. (2016). Burden of disease and costs of exposure to endocrine disrupting chemicals in the European Union: An updated analysis. *Andrology* 4, 565-572. <https://doi.org/10.1111/andr.12178>
- Tung, C-W. (2015). ChemDIS: A chemical-disease inference system based on chemical-protein interactions. *J Cheminformatics* 7, 25. <https://doi.org/10.1186/s13321-015-0077-3>.
- U.S. EPA – U.S. Environmental Protection Agency (1999). Integrated Risk Information System (IRIS) on Hexachlorobenzene. National Center for Environmental Assessment, Office of Research and Development, Washington, DC.
- Yang, W-S., Li, H-L., Xu, H-L. et al. (2016). Type 2 diabetes and the risk of non-Hodgkin's lymphoma: A report from two population-based cohort studies in China. *Eur J Cancer Prev* 25, 149-154. <https://doi.org/10.1097/CEJ.0000000000000150>

Conflict of interest

The authors declare they have no actual or potential competing financial interests.

Acknowledgments

We would like to acknowledge EU-ToxRisk, a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement No 681002.

Correspondence to

Karine Audouze, PhD
Inserm UMR-S 973
Molécules Thérapeutiques in silico
Paris Diderot University
35 rue Hélène Brion
75013 Paris, France
e-mail: karine.audouze@univ-paris-diderot.fr