ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201

Research Article

Human Environmental Disease Network: A computational model to assess toxicology of contaminants

Olivier Taboureau[1,2,3] and Karine Audouze[1,2]
[1]INSERM UMR-S973, Molécules Thérapeutiques in silico, Paris, France; [2] University of Paris Diderot, Paris, France; [3]Novo 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 for diverse human disorders. However, the relationships between diseases based on chemical exposure have been rarely 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 on systems biology and chemical toxicology using chemical contaminants information and their disease relationships from the reported TDDB database.
The resulting human EDN takes into consideration the level of evidence of the toxicant-disease relationships allowing including some degrees of significance in the disease-disease associations. Such network can be used to identify uncharacterized connections between diseases. Examples are discussed with type 2 diabetes (T2D). Additionally, this computational model allows to confirm already know chemical-disease links (e.g. bisphenol A and behavioral disorders) and also to reveal unexpected associations between chemicals and diseases (e.g. chlordane and olfactory alteration), thus predicting which chemicals may be risk factors to human health.
With the proposed human EDN model, it is possible to explore common biological mechanism between two diseases through chemical exposure helping us to gain insight into disease etiology and comorbidity. Such 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, our understanding about disease-causing defects is still a challenge(Hunter, 2005). The emergence of one or more additional disorders co-occurring with a primary disease, i.e. comorbidity, due to a given treatment, lifestyle or environmental exposure, give a higher degree of complexity.

One problem is to identify how diseases are connected to each other within the aim to better understand the etiology and pathogenesis of similar diseases. Several networks based approaches have allowed deciphering diseases 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 together (Goh et al., 2007). The network consisted 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.* created an approach for gathering phenotypic descriptions of patients from medical records suggesting new disease-disease associations (Roque et al., 2011). Although all these studies provided comprehensive views of links between diseases, they rely on existing knowledge, i.e. gene, pathways and phenotypic associations (Hidalgo et al., 2009).

Epidemiological and biological studies have suggested that a number of environmental chemicals may play causative role for some human disorders.Nevertheless, human are potentially exposed to more than 80,000 substances for which toxicity remains largely unknown. 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 (National Research Council 2007)have recommended that *in silico*

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

approachesshould be included in future assessments of toxicity within the aim to reduce and refine existing methods. While *in silico* computer simulation cannot substitute *in vitro* or *in vivo* testing, they can help focuson particular substances and targets to allow priority setting for more efficient testing. Overall computational approaches can be an alternative to reduce the animal experiments according to the 3R definition (Replacement, Reduction, Refinement).

Computational systems biology studies have shown the possibility to link chemicals to diseases such as the pesticide dichlorodiphenyltrichloroethane (DDT) and type II diabetes (Audouze and Grandjean, 2011),and persistent organic pollutants (POPs) and metabolic diseases (Ruiz et al., 2015). However, this area has not been screened systematically and environmental factors are rarely considered when creating disease-disease networks. Some investigations have been reported on this area. A chemical-disease inference system based on chemical-protein interactions, ChemDIS, has been performed within the aim to predict potential health risk associated to chemicals (Tung, 2015). Environmental etiological factors and genetic factors associations using Medical Subject Headings (MeSH) terms have been described (Liu et al., 2009).Finally, 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 limited usually 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)(http://www.healthandenvironment.org/tddb). This database compiled existing information between chemical contaminants and approximately 200 human diseases based on biological and epidemiological evidences. Interestingly, they included a level of evidence (from limited to good) between chemicals and diseases to estimate how the chemicals exposure could contribute to environmental diseases. For example, mercury is associated to cognitive impairment and behavioral problems with high level of evidence whereas methoxychlor is associated to reduce male fertility with limited evidence. To develop our EDN, we took advantage of the computational network biology approach that we developed with toxicogenomics data (Audouze et al., 2010). The concept originally based on Protein-Protein Associations Network (P-PAN) can be transposed in other area and has shown some success with experimental validation of identification of novels associations between chemicals and biological targets (Audouze et al., 2014). With the TDDB database, although, the list of chemicals and diseases associated is far from being complete, the database has the advantage to have 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). We will present a case study to demonstrate the ability of our EDN model to predict disease-disease connections, chemical-disease links, and we will propose potential new associations. To our knowledge, this is the first time that a disease comorbidity network based on environmental chemicals is 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 for 2790 connections between 601 environmental chemicals and 198 human diseases. We used the three available strengths of evidence to create a three levels disease network: The ‘strong evidence’ (SE) represents verified link between chemical and disease. Three cases are 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 cohorts studies, and (c) the chemicals are listed as group 1 human carcinogens by the International Agency for Research on Cancer (IARC). The ‘good evidence’ (GE), which include chemical-disease associations based on epidemiological studies, chemicals listed in the IARC group 2A (limited evidence for humans and strong for animal) and chemicals listed by OEHHA´s Prop 65 program. Finally, the ‘limited evidence’ (LE) is the last category. It contains chemicals associated to diseases based on case reports, on conflicting evidence and chemicals listed in the IARC group 2B and the EPA group B2 (limited in humans and in animals). Although the extracted data may be limited, the TDDB database represents the most complete repository of chemicals associated to human diseases with evidence information.

Generating a high confidence human disease-disease network

The relevant chemical-disease links collected from the TDDB database were used to generate the disease-disease network. Taking all the information, the maximum number of diseases associated to chemicals i.e. pesticides is 99, and the maximum number of chemicals associated to one disease i.e.fetotoxicity is 70. Looking only at the strong evidence layer, the maximum number of diseases (17 diseases)is linked to lead, and the maximum number of chemicals (17 chemicals) is associated to the hepatitis disease. The EDN was created by instantiating a node for each disease, and linking by an edge any disease-disease pair where at least one overlapping chemical was identified. 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, … D_n.

ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201

For each disease, a set of chemicals is associated:
$Chem_x$:= { $c \in$ Chemicals|c is associated with D_x }
This probabilistic score (pS) between a pair of diseases is calculated by the following equation:

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

The higher is the pS score, the strongest is the association.

Exploration of the biological mechanisms: biological enrichment
To identify biological outcomes potentially related to individual chemical, 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 (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 chemicalusing a statistical test based on a hypergeometric distribution. ABonferroni correction for multiple testing of *p*-values was used to select the most relevant associations.

Predicting novel chemical-disease links
To predict diseases potentially linked to a chemical, a neighbor disease approachwas performed based on the neighbor protein procedure previously described by Audouze et al. (Audouze et al., 2010). This approach is a multi-steps procedure. Firstly, diseases associated to the chemicals of interest from TDDB database are listed. Then, these input diseases allow identifying network(s) surrounding them by using 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. For each neighbor, a score was calculated 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 in the conservative end of the optimal range for disease-disease interaction networks. Finally, all diseases in the network(s) were checked for associations among them, and the missing one were added. A confidence score was established by testing each network for enrichment on the input set by comparing them against 1.0e-4 random networks. The individual disease score was used to rank them, allowing predicting potential diseases linked to the chemical.

3 Results

3.1 Generating an environmental disease network (EDN)
Based on chemical-disease associations extracted from the TDDB database, we have constructed a human EDN using the three levels of evidence i.e. SE, GE and LE. An overview of the strategy is shown on Figure 1. In total, the resulting EDN consists of 6258 associations between 196 diseases. The SE level contains 125 diseases interconnected, 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) to each disease-disease association represented by the weight of each link (see Methods).

3.2 Mining the environmental disease network
For an easier interpretation, each disease isclassified into 19 primary disorder classes following the classification scheme described in the human disease network (Goh et al., 2007) (see Tab. S1). 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 represent themetabolism class. Interaction of two diseases belonging to the same class is defined as intra-class.Two diseases from two different classes is defined as inter-class. When a disease affect several biological systems, only one class is assigned to the disease, based on the systems know to be the most affected. Then each disease was annotated to only one of the 19 classes.
All together, the EDN displays connections between 196 diseases intra- and inter- classes. 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 mechanism of actions 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, excepted cancer disorders, which are sparse in the network. Not surprising, among the top intra-class associations, some well-known diseases are retrieved (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. 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 (see Figure S1) and LE (see Figure S2) 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-classes 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 previous studies, where 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-classes, and most of them concern the cancer's one (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 potentially 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.

3.3 Case study: Type II diabetes

EDN exploration: prediction of T2D-disease associations

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

From the LE layer (weaker evidences) we can see associations between T2D and other disorders from the reproductive system and hypertension. Associations between T2D and diseases from the reproductive system such as abnormal sperm and reduced male fertility were linked via polychlorinated biphenyl compounds (PCBs). Previous studies have shown that diabetes may affect male reproductive functions at multiple levels including diminished sperm quality (Jangir and Jain, 2014). These hypotheses are also well in line with 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-value of 0.043). From the ChemDis webserver(Tung, 2015), the association between PCB 126 and diabetes mellitus is statistically inferred from a chemical-protein-disease relationship (p-value of 1.98^{e-2}).

About the second association, T2D-hypertension, previous studies have supported a substantial overlap between diabetes and hypertension in etiology and disease mechanisms (Gress et al., 2000;Cheung and Li, 2012). This interaction has been explained by insulin treatmentfor diabetic patients who presented some hypertension symptoms(Jensen et al., 2012). In our approach, both diseases are connected through the exposure to PCBs. Looking through the literature, it has been reported independently that PCBs might increase insulin resistance, cause diabetes(Kouznetsova et al., 2007) and are associated to hypertension (Everett et al., 2008). However, the mechanism(s) remains to be ascertained. Exploring HExpoCHem, the PCB 126 was associated to insulin sensitivity via a protein complex of ADIPOQ (p-value non significant), and to hypertension arterial via a protein complex of AGTR1 and another complex of NOS3(p-values non significant). From 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 relation in the mechanism of action.

In the two others layers (SE and GE), fewer associations were identified. For example, a link between T2D and Hodgkin's disease (HD) is assumed via endocrine disruptor chemicals in GE (Figure 3). Such association is supported by a study, which has examined epidemiological associations between diabetes and the risk of HD, and concluded that these diseases may be linked (Mitri et al., 2008). Considering the significant increase of T2D and HD during the past decades, a study on a population-based cohort of more than 130 000 adults suggests that T2D might be associated with an increased risk of 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 changes via 2,3,7,8-tetrachlorodibenzo-p-dioxin (TCDD) (GE layer) and dichlorodiphenyltrichloroethane (DDT) (LE layer). Recent computational studies have explored possible pathogenetic linkage between environmental chemicals and diseases through various data types such as genome-wide associations and disease similarities, and found a potential link between T2D and TCDD and, T2D and DDT (Audouze and Grandjean, 2011).

The EDN shows also that dermatological disorders might 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 have a better comprehension of the predicted T2D-hormonal changes associations 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, given a significant adjusted p-value of 1.904^{e-19}. From the GO process, among the 14,650 genes, 162 were linked to regulation of hormone secretion. 14 among them are associated to TCDD given a p-value of 3.361^{e-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 diabetes, 16 genes associated to DDT were know 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-value of 1.596^{e-05}) 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 novel chemical-disease risks

Besides revealing connections between diseases, the EDN can be used to assess the risk of a chemical to induce diseases. It is now well define that chemical does not interfere with a unique protein target but may interact with several proteins (Paolini et al., 2006). As chemicals are linked to several proteins or protein complexes, and that diseases are also connected to multi-proteins, it is essential to identify potential relationships between chemicals and diseases within the aim to have a better biological understanding. Based on this assumption and in order to suggest novel connections between chemicals and diseases, we developed and applied to the EDN model a neighbor disease procedure built on a neighbor protein procedure, which score the associations between diseases (details in Methods). The performance of this approach has been shown in previous studies(Audouze et al., 2010;Audouze et al., 2014). To suggest chemical-disease risks, only the high significant evidence (SE) layer was used. Diseases known to be connected to various chemicals were listed independently i.e. one diseases 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 systems, 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). By today no direct epidemiological associations exist, meaning that it is difficult to say that BPA causes cancer growth, but it could well contribute as it is disrupting the genes that defend against such growth (Bhan et al., 2014).

A relation between BPA and behavioral disorders is also depicted in the network. From the literature, 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). 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 also on chlordane as a second example. Chlordane is a pesticide banned in Europe and US, but very persistent in the environment. Therefore exposure to chlordane is still harming the health of millions of people. The acute (short-term) effects of chlordane in humans consist of gastrointestinal distress and neurological symptoms, such as tremors and convulsions.Chronic (long-term) inhalation exposure of humans to chlordane results in effects on the nervous system (United States Environmental Protection Agency: U.S. EPA). Using our approach, chlordane is predicted to have an action on brain cancer, which is in line with the US EPA, who classified chlordane as a Group B2, probable human carcinogen.Chlordane was also predicted to be associated to olfactory alteration prediction. 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 produce such damage (Doty, 2015). Furthermore, it has been suggested that olfactory lost can occur as a result of exposure to chemicals present in the air pollution or workplace situations(Quandt et al., 2016), but no specific link between chlordane and olfactory disorders have been yet reported.

We finally explored, the synthetic industrial chemical hexachlorobenzene (HCB). HCB is a bioaccumulative, persistent, and toxic pollutant. Historically HCB was commonly used as a pesticide and fungicide. Although HCB production has stopped in many countries, the compound is still generated inadvertently, as a byproduct and/or impurity in the manufacture of various chlorinated compounds, and released into the environment. In our network, HCB is predicted to be associated to reproductive disorders as congenital malformation and abnormal sperm. In the literature, environmental exposures 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(Reed et al., 2007).

4 Discussion

The proposed approach offers a network-based hypothesis for the emergence of complex diseases, which cannot always be explained by genetic.

Although, our EDN can be of help in the understanding of disease co-occurrences, of mechanism of action,and in the risk assessment of new chemicals, we are aware that the chemical-disease annotations used in this workis limited in term of diseases and chemicals. For example, the used version of the TDBB database 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-prediction and an extension of the proposed network with integration ofothers data would be beneficial. Some efforts are on going in

measuring potential human exposure to environmental chemicals and to facilitate the public access to these data. As example, we can notice the exposure data collection from the U.S.EPA agency, the human exposome project (http://humanexposomeproject.com/), the Heals project that is the largest research project in Europe on environment and health (http://www.heals-eu.eu/)and the national report on human exposure to environmental chemicals by the centers for disease control and prevention (https://www.cdc.gov/exposurereport/).

Still, the TDBB database has the originality to organize 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. In general a bipartite network is considering only two levels of information (yes/no, inhibitor/activator) without considering the degree of evidence.

Developing biological networks with more selective data allow 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 newsuitable databases for the generation of computational methods able to decipher potential risks associated to chemicals and to generate hypothesis, accelerating the hazard identification process. Within the aim to screen as much as possible of available data, one way would be to use advanced text miming tools, such as the one used to extract drug-adverse event information from electronic medical records(Roitmann et al., 2014). Crossing the hypothesis made by our approach with some other observations made in the literature would be also valuable to characterize potential chemical-disease relationships. For example, a group lead by Trasande L. has recently developed a system to estimate health and economic costs related to endocrine disrupting chemicals (EDC) exposure in the European Union(Trasande et al., 2015, 2016). To estimate costs they used available epidemiological and toxicological evidence for each EDC and weighted them.Therefore such probability of causation, e.g. EDC causation for IQ loss and association with autism, childhood obesity or male infertilitycould be crossed with other computational models and the TDDB database.

5 Conclusion

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-disease associations was illustrated by several examples, as the suggestion of links between BPA and ADHD; and chlordane and olfactory dysfunction thatwould need further investigations to be confirmed.Such computational method can be used to enhance our current knowledge, and may help for 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. http://dx.doi.org/10.1371/journal.pone.0150762.

Audouze, K. and Grandjean, P. (2011). Application of computational systems biology to explore environmental toxicity hazards. *Environ. Health Perspect*. 119:1754–1759.http://dx.doi.org/10.1289/ehp.1103533.

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. http://dx.doi.org/10.1371/journal.pcbi.1000788.

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. http://dx.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. http://dx.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.*141:160-70. http://dx.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. http://dx.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. http://dx.doi.org/10.1007/s11883-012-0227-2.

Davis, D. A., Chawla, N. V. (2011). Exploring and exploiting disease interactions from multi-relational gene and phenotype networks. *PloS One* 6:e22670. http://dx.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(5710):724-7. http://dx.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. http://dx.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.http://dx.doi.org/10.1016/j.neuro.2012.05.010

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. http://dx.doi.org/10.1016/j.envres.2008.05.006

Gene Ontology Consortium. (2015). Gene Ontology Consortium: going forward. *Nucleic Acids Res.* 43:D1049–1056. http://dx.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. http://dx.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. http://dx.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.http://dx.doi.org/10.1056/NEJM200003303421301

Grondin, C. J., Davis, A. P., Wiegers, T. C. et al. (2016). Advancing Exposure Science through Chemical Data Curation and Integration in the Comparative Toxicogenomics Database. *Environ. Health Perspect.* Epub ahead of print. http://dx.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. http://dx.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. http://dx.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. http://dx.doi.org/10.1371/journal.pcbi.1000353.

Hunter, D. J. (2005). Gene-environment interactions in human diseases. *Nat. Rev. Genet.* 6:287–298.http://dx.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.http://dx.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.http://dx.doi.org/10.1038/nrg3208.

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.http://dx.doi.org/10.1093/nar/gks1166.

Knudsen, T., Martin, M., Chandler, K. et al. (2013). Predictive models and computational toxicology. *Methods Mol Biol*.947: 343-74. http://dx.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(1): 11-22. http://dx.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. http://dx.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. http://dx.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.http://dx.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.http://dx.doi.org/10.1186/1471-2105-10-S2-S14.

National Research Council (2007). Toxicity Testing in the 21st Century: A Vision and a Strategy. Washington, DC: National Academy Press.

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. http://dx.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.http://dx.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.http://dx.doi.org/10.2337/dc08-1034.

Mustieles, V., Pérez-Lobato, R., Olea, N., Fernández, M. F. (2015). Bisphenol A: Human exposure and neurobehavior. *Neurotoxicology* 49:174–184.http://dx.doi.org/10.1016/j.neuro.2015.06.002.

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. http://dx.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. http://dx.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* 53:12–1.http://dx.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. Am. Coll. Occup. Environ. Med*. 58:248–253.http://dx.doi.org/10.1097/JOM.0000000000000672.

Reed, L., Buchner, V. and Tchounwou, P. B. (2007). Environmental toxicology and health effects associated with hexachlorobenzene exposure. *Rev. Environ. Health* 22: 213–243. http://dx.doi.org/10.1515/REVEH.2007.22.3.213

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. http://dx.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. http://dx.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. http://dx.doi.org/10.1371/journal.pcbi.1002141.

Ruiz, P., Perlina, A., Mumtaz, M. and Fowler, B. A. (2015). A Systems Biology Approach Reveals Converging Molecular Mechanisms that Link Different POPs to Common Metabolic Diseases. Environ. Health Perspect. Epub ahead of print. http://dx.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.http://dx.doi.org/10.1093/database/baq020.

Sun, K., Buchan, N., Larminie, C. and Pržulj, N. (2014a). The integrated disease network. *Integr. Biol. Quant. Biosci. Nano Macro* 6:1069–1079. http://dx.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.http://dx.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.http://dx.doi.org/10.1371/journal.pcbi.1000662.

Taboureau, O., Jacobsen, U. P., Kalhauge, C. et al. (2013). HExpoChem: a systems biology resource to explore human exposure to chemicals. *Bioinformatics*29:1231–1232. http://dx.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.http://dx.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* 4:565-72. http://dx.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. http://dx.doi.org/10.1186/s13321-015-0077-3.

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. Off. J. Eur. Cancer Prev. Organ. ECP* 25:149–154.http://dx.doi.org/10.1097/CEJ.0000000000000150.

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.

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

Correspondence to
KarineAudouze, PhD
Inserm UMR-S973
MoléculesThérapeutiques in silico
Paris Diderot University
35 rue Hélène Brion
75013 Paris, France.
e-mail: karine.audouze@univ-paris-diderot.fr

ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201

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	Nb. gene	P-val.	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 function	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. indicates corrected p-values non significant (>1).

Tab.2: Mining the full EDN for diseases associated to bisphenol A (BPA), chlordane and hexachlorobenzene (HCB). The number of diseases already known from the database to be associated to 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, bisphenol A is predicted to be connected to eight already known diseases, for which menstrual disorders as the highest score. Moreover, BPA is also predicted to be linked to five other diseases, among them leukemia; which has the highest score, and no literature support by today.

Environmental exposure	Known	diseases	Predicteddiseases	Score
	(all layers)	(max and min score)		
Bisphenol A	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)	Braincancer 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

ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201

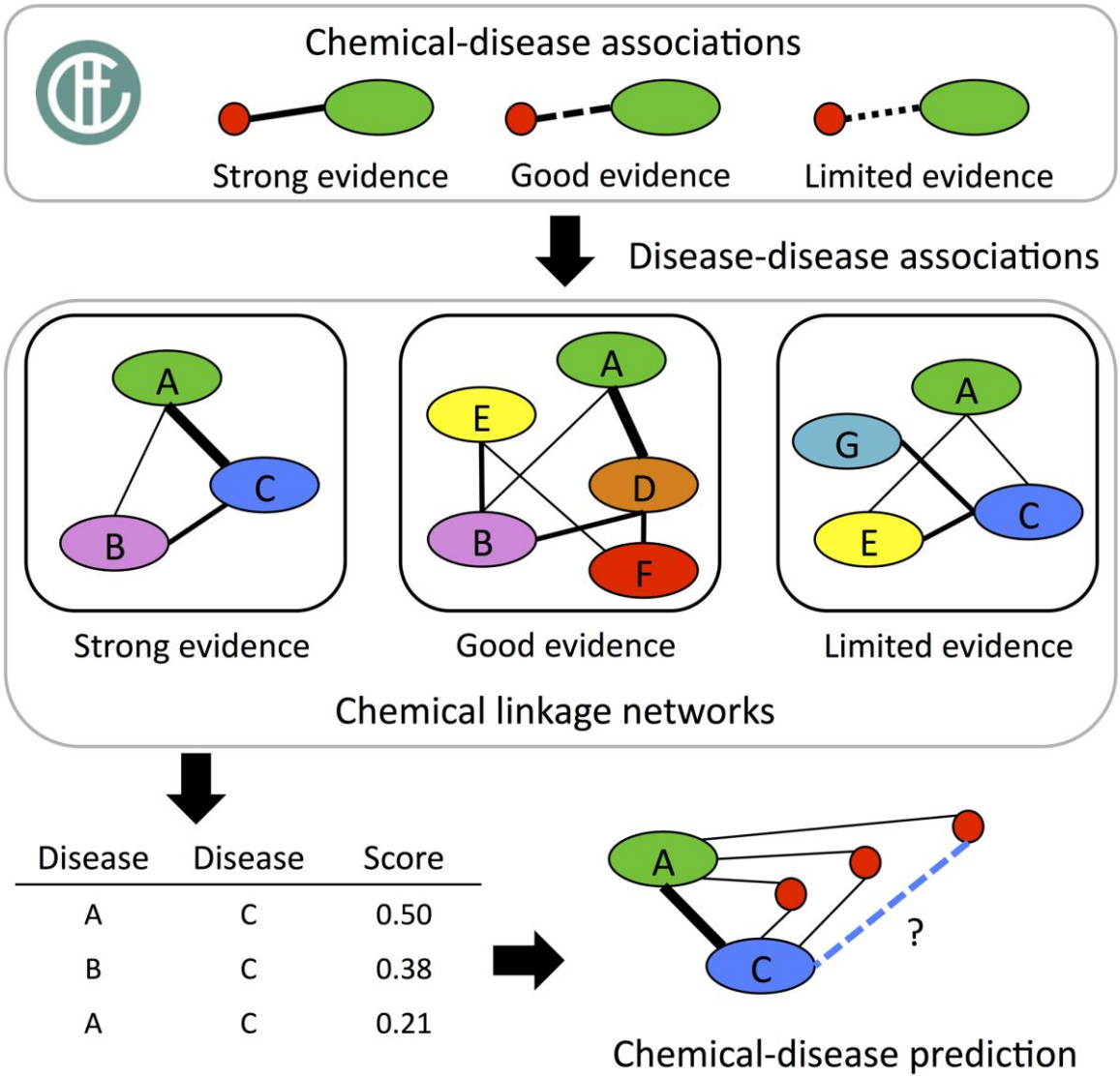


Fig. 1: Workflow of the proposed systems chemical biology strategy for predicting disease-disease and chemical-disease associations.
Information on chemicals known to be linked to diseases, and their evidence levels, are extracted from the TDDB database and cleaned. The EDN model was then created using these data based on a protein-protein association network procedure, assuming that two nodes (i.e. diseases) are connected to each other if they shared at least one chemical for which a causal evidence is associated with both diseases. A probabilistic score was assigned to each disease pair in order to rank them. The higher is the score, the stronger is the association. Using a network-neighbor's pull down procedure, prediction of connection between chemical and disease was performed.

ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201

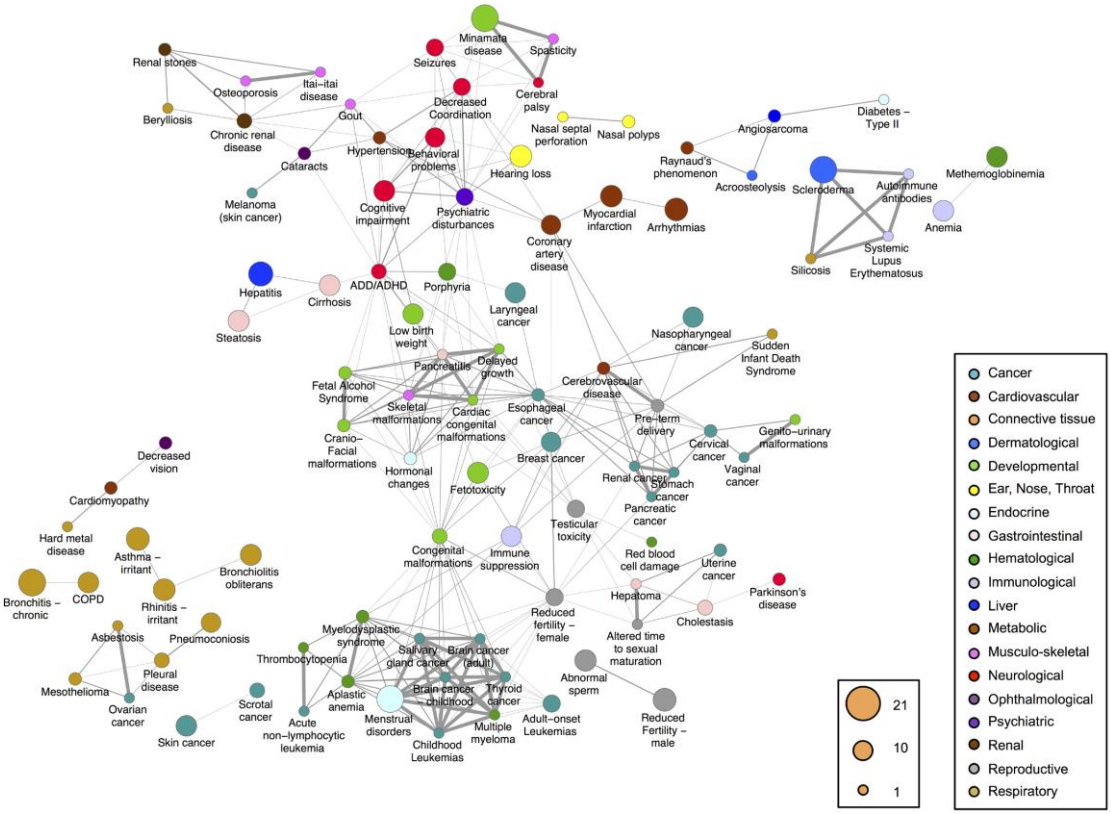
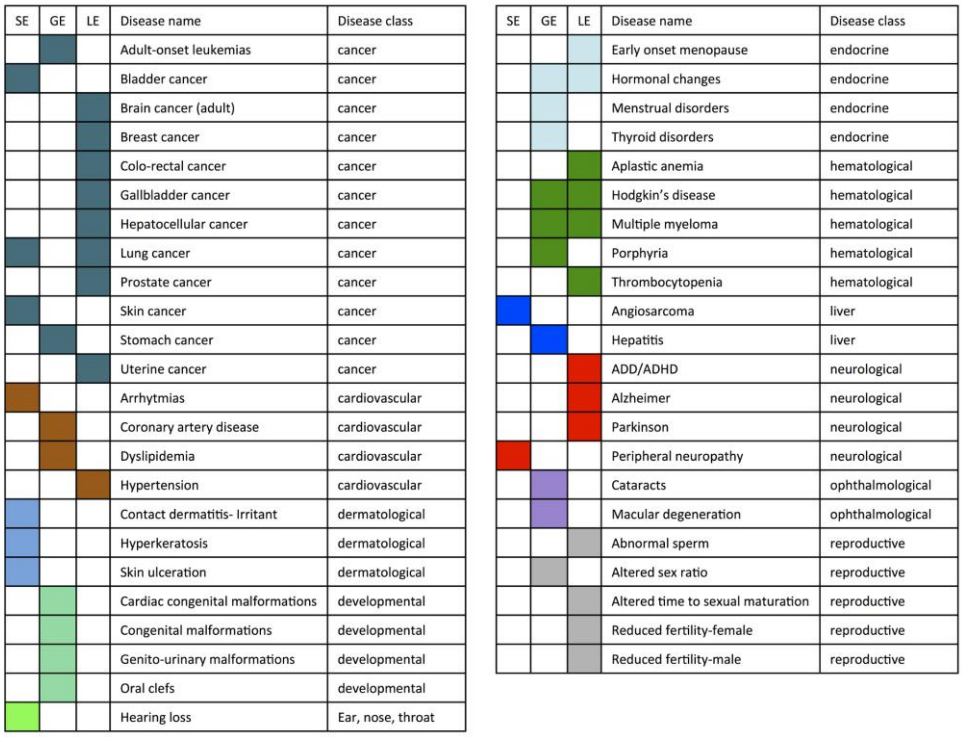


Fig. 2: Representation of the top significant disease-disease associations based on the probabilistic score using the strong level of evidence (SE).
Each node corresponds to a unique disease, colored according to the biological system (class) to which it belongs. The name of the 19 systems is shown on the right. Node's sizes are according to the number of chemicals linked to the disease with a strong evidence level. An edge is placed between two diseases if they share at least one chemical within the SE level. The width of a link is proportional to the number of chemicals that are linked to both diseases. For example, six chemicals are linked in both myocardial infarction and arrhythmias disorders, resulting in an edge with a probabilistic score of 0.26. Only the top significant associations are shown (based on the calculated score) for clarity of the figure.

ALTEX Online first
published October 21, 2016
https://doi.org/10.14573/altex.1607201



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
			Arrhytmias	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 clefs	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 allowing to see on which class belong a disease.