

Integrative systems toxicology to predict human biological systems affected by exposure to environmental chemicals.

Olivier Taboureau^{a*}, Walid El M'Selmi^a, Karine Audouze^{b*}

^a Université de Paris, BFA, CNRS UMR 8251, ERL Inserm U1133, F-75013, Paris, France

^b Université de Paris, T3S, Inserm UMR-S 1124, F-75006, Paris, France

*corresponding authors: karine.audouze@u-paris.fr & olivier.taboureau@u-paris.fr

Abstract

Biological systems are disturbed by several factors that are defined by the exposome. Environmental substances, including endocrine disruptors (EDs), represent the chemical exposome. These stressors may alter biological systems, that could lead to toxic health effects. Even if scientific evidence provide links between diverse environmental substances and disorders, innovative approaches, including alternative methods to animal testing, are still needed to address the complexity of the chemical mechanisms of action. Network science appears to be a valuable approach for helping to decipher a comprehensive assessment of the chemical exposome. A computational protein system-system association network (pS-SAN), based on various data sources such as chemical-protein interactions, chemical-system links, and protein-tissue associations was developed. The integrative systems toxicological model was applied to three EDs, to predict potential biological systems they may perturb. The results revealed that several systems may be disturbed by theses EDs, such as the kidney, liver and endocrine systems. The presented network-based approach

24 highlights an opportunity to shift the paradigm of chemical risk assessment towards a better
25 understanding of chemical toxicology mechanisms.

26

27 **Keywords:** chemical exposome, endocrine-disrupting chemical, integrative systems biology,
28 computational model, network science, new approaches methodologies (NAMs), OBERON.

29

30

Introduction

Through their entire life, humans are exposed to various hazardous substances from diverse sources such as their environment, diet or medical treatment to cite but a few sources. All of them represent the chemical part of the exposome (Wild, 2005). The interplay between genes, lifestyle, chemical exposure and disease etiology is now well accepted (Judson et al., 2012), and integration of the genome and the exposome is a novel paradigm in toxicology (Barouki et al., 2018)(Vermeulen et al., 2020). Evidence on the impact of some of these chemicals on human health have considerably increased over the last years. Notably several health outcomes, such as cancers, metabolic disorders, neurocognitive functions, infertility, immune diseases, and allergies have been associated with chemical exposure such as endocrine disruptors (EDs) and persistent organic pollutant (POPs) (Landrigan et al., 2018) (Agier et al., 2019) (Wu et al., 2020). Even if some mechanisms of action have been identified by biological experiments and computational predictive models, there is still a need to develop integrative test methods, which include *in vitro*, non-vertebrate *in vivo*, and *in silico* models to detect EDs (Audouze et al., 2020). In the context of reduction, refinement and replacement of animal use (3Rs), more and more European and US directives support the move from traditional animal models to new approach methodologies (NAMs) in chemical risk assessment (Bopp et al., 2019). Such alternative methods, including integrated approaches to testing and assessment (IATAs) allow evaluating large numbers of uncharacterized substances including EDs, and also reducing time and cost of current approaches (Sakuratani et al., 2018)(Audouze et al., 2020). One of the most challenging problems in biomedical research and chemical risk assessment is to understand the underlying mechanisms of complex diseases, and several programs have been launched on these purposes. For example, the OECD has recently developed the concept of adverse

outcome pathway (AOP) to support toxicity evidence with mechanistic pathways and mode of action for chemical safety and risk assessment (Ankley et al., 2010)(Leist et al., 2017). Also, the National Research Council in US outlined a general strategy for non-animal testing approaches: Toxicity testing in the 21th Century in 2007 (Tox21)(Krewski et al., 2010), which have recommended to include, among others, computational toxicology and *in silico* approaches in future assessments of toxicity as an inexpensive and efficient tool for screening purposes (Knudsen et al., 2013). While computational studies cannot substitute for *in vitro* or *in vivo* testing, they can help in the assessment of human health risks and prioritization of chemicals. Computational systems biology models allow identification of linkages between chemicals and diseases. For example, associations of *in vitro* assays and AOPs to test chemicals with potential toxicological effects have been reported, as well as computational approaches compiling several and different sources of information(Knapen et al., 2015) (Browne et al., 2017) (Pittman et al., 2018)(Aguayo-Orozco et al., 2019)(Carvaille et al., 2019)(Rugard et al., 2020)(Jornod et al., 2020). One limitation of these models is that they are specific to adverse outcome, cell line, or available knowledge, and not on an entire tissue or system. A recent study, showed that even if genes carrying mutations associated with genetic diseases are present in all human cells, clinical manifestations of such diseases are tissues-specific, and need a whole functional subnetwork of genes to be expressed in that tissue (Kitsak et al., 2016). We propose an innovative computational strategy to predict chemical effect at the human systems level. Based on the integration of chemical-toxicity effects, chemical-protein interactions, protein-tissue expressions and tissue-toxicity relations, an integrative model was developed to predict systems targeted by chemicals that are highly associated to adverse outcomes. As a proof of concept, the developed p-SSAN

78 model was applied to three compounds, acetaminophen, valproic acid and 2,3,7,8-
79 tetrachlorodibenzo-p-dioxin (TCDD), suspected to have endocrine-disrupting properties.

Materials and Methods

Chemical-system associations

We extracted data from the Registry of Toxic Effects of Chemical Substances (RTECS) from Biova (<http://accelrys.com/products/collaborative-science/databases/bioactivity-databases/rtecs.html>). The RTECS database is a comprehensive resource of toxicity information, compiling known and available data (including affected systems) for environmental chemicals (more than 174,000 chemical substances) such as drugs, food additives, pesticides, fungicides, plastics, household products, etc. In the present study, a biological system is defined by a code from the RTECS database classification (e.g. the reproductive system that includes 'effects on embryo or fetus'; 'effects on fertility') (Table 1). Only human data related information was considered.

To avoid problem due to multiple chemical names and synonyms, and to facilitate further data analysis, chemicals were annotated with a RTECS identifier, and a canonical SMILES code (Simplified Molecular-Input Line-Entry System) characterizing the structure of a substance. All canonical SMILES were converted into IUPAC International Chemical Identifier (InChIKey), a standard and universal way to encode chemicals. Finally, for each unique compound, the PubChem identifier (CID) was kept as final chemical identifier. All different steps (and the following ones) were performed using scripts in python.

Chemical-protein associations

The available chemical-protein associations knowledge were extracted using the STITCH database (version 5) (Babin and Gibbons, 2009). The STITCH database contains integrated information from various disparate data sources. It combines interactions between 430,000 chemicals and proteins, for various living organisms.

To capture the most reliable information, only interactions that reached at least a confidence score >0.4 were extracted, which is the medium confidence score according to STITCH (in a range between 0 and 1). Moreover, only interactions involving human proteins were kept. All chemicals extracted from STITCH were also encoded with an InChiKey, and annotated to CID in order to facilitate the chemicals mapping from STITCH to RTCES. To facilitate data integration, proteins were mapped to their corresponding Ensembl ID (ENSP ID).

Protein-system associations

First, association between proteins and systems was performed using the previously compiled chemical-system and chemical-protein information, using the CID identifier of the chemicals. In a second step, to mimic the true biological organization and function, i.e. all proteins are not expressed in all systems, we went one step further to keep only the relevant information and avoid false positive. To filter the data, we used the Human Protein Atlas (HPA) database (<https://www.proteinatlas.org>) (Thul et al., 2017). The HPA database is divided in three sub-atlas, that are the tissue atlas, the pathology atlas, and the cell atlas. For our study, only the tissue atlas was considered. The tissue atlas contains information related to 76 different cell types, corresponding to 44 non-disease human tissue types covering all major parts of the human body, and protein expression data covering 15,297 of the protein coding genes. It includes data for genes and their corresponding Ensembl gene annotations (ENSG ID), tissues and their relationships at several levels of expression (non-detected, uncertain, low, medium, high). Only information for normal tissues with gene expression levels 'medium' and 'high' were kept in our study. Then, the 24 systems from RTCES were manually mapped to the 55 tissues from the HPA database (for example, 'epididymis' and

'testis' were mapped to the 'reproductive' system). For some systems, such as 'biochemical' or 'chronic data' it was not possible to map them to tissues. Therefore, we conserved these systems separately. The relations between tissues and systems are described in Table 2. Overall, the mapping of chemical-protein-system from the different sources of data represents the foundation of our integrative toxicological systems network.

High confidence human system-system association network models

Using the chemical-system and protein-system information collected and matched from the various data sources, two human system-system association based-network models (S-SAN) were developed. A first model using the chemical information, named cS-SAN (chemical system-system association network), and a second model using the protein information, called pS-SAN (protein system-system association network). To generate high confidence models, a previously described computational network biology approach was used (Audouze et al., 2010). In both S-SAN models, each system is depicted by a node, and links between them are represented by an edge. Edges represent any system-system pairs where at least one shared chemical (cS-SAN) or protein (pS-SAN) was identified. For both models, non-redundant lists of information were verified, meaning that if systems X and Z were connected, two possibilities may be present into the network (X-Z and Z-X). Only one of these associations was kept to create the S-SANs.

Probabilistic score

To reduce noise, and select the most significant system-system associations in both models, we assigned a probabilistic score (pS) to each generated system-system pair, based from previous experience (Audouze et al., 2010; Taboureau & Audouze, 2016). This score is based

on the probability that a chemical or a protein linked to a system A will also affect the other system B. The systems are represented by $S_1, S_2, \dots S_n$. For each system, a set of chemicals or proteins is associated:

$\text{Chem}_x := \{ c \in \text{Chemicals} \mid c \text{ is associated with } S_x \}$

$\text{Prot}_x := \{ c \in \text{Proteins} \mid c \text{ is associated with } S_x \}$

This probabilistic score (pS) between a pair of systems is calculated by the following equations (using the chemicals for the cS-SAN and the proteins for the pS-SAN models:

$$\text{pS}(S_a, S_b) = \frac{|\text{Chem}_a \cap \text{Chem}_b|}{|\text{Chem}_a \cup \text{Chem}_b|} \quad \text{pS}(S_a, S_b) = \frac{|\text{Prot}_a \cap \text{Prot}_b|}{|\text{Prot}_a \cup \text{Prot}_b|}$$

A higher pS score indicates a stronger association.

Prediction of human systems affected by endocrine-disrupting chemicals

Chemical data: Three widely used chemicals were selected due to their potential endocrine disrupting activities, and tested on the developed pS-SAN model. Acetaminophen (AC), also known as paracetamol, is an analgesic used to treat pain and fever (Konkel, 2018). Studies have shown that mild analgesics could have an impact on human health, and may cause multiple endocrine disturbances in the male reproductive systems via for example the inhibition of the prostaglandin pathways (Kristensen et al., 2011). Valproic acid (VA), is a seizures and antiepileptic drug that has been shown to affect the metabolic and endocrine systems resulting in hormonal disturbances and incidence of some metabolic diseases. Among its serious known adverse effects, VA has been associated to human abnormalities such as hypospadias, and body weight gain (Raghavan et al., 2018)(Verrotti et al., 2011). TCDD is a dioxin, well known as a contaminant of agent orange, a herbicide that was among other, released during the Seveso disaster. TCDD have been demonstrated to alter the

endocrine immune and the nervous systems. Exposure to TCDD may lead to various adverse effects on human, linked to for example to reproduction, development, hepatotoxicity, diabetes and cancer (Yoshioka and Tohyama, 2019).

Prediction using protein complexes

To predict which system(s) can be affected by these three EDs, known chemical-protein links were extracted from the Comparative Toxicogenomics Database (CTD) (Davis et al., 2017). The CTD is a database providing manually curated information regarding chemical gene/protein associations (1,518,440 unique associations for all organisms covered). Only information regarding curated data and human proteins were retained. Then, to assess systems potentially affected by a chemical, over-representation analysis (ORA) were performed. Using the developed protein system-system network (pS-SAN), statistical significance based on hypergeometric distribution was calculated individually for each of the three EDs, using their protein lists extracted from the CTD database. A significance level of 0.05 after Bonferroni correction for multiple testing of the p-values was used to select the most relevant associations (Lee and Lee, 2018). As results, systems were associated to each of the three EDs, via their protein lists. Only the single most significant system associated to each ED was reported. The workflow of the approach is depicted in Figure 1.

Results and discussion

System-system association network in the chemical space (cS-SAN)

From the RTECS database, toxicity information was extracted for 52,625 chemicals that are annotated to 24 systems (Table 1), with a total of 91,846 associations. We noticed that many chemicals are annotated to more than one toxicological system. Therefore, to have an overview, and to assess the systems highly impacted by those chemicals, a chemical system-system associations network (cS-SAN) was developed, using the 24 human systems and all chemicals reported in RTECS. The resulting cS-SAN appeared to be a complex graph containing a total of 271 unique associations between the 24 systems (see details Table S1). Based on the pS score, only 16 systems, among the 24 systems, had high significant associations ($pS > 0.07$) via 27 connections (Figure 2). The maximum of shared compounds was 5434, between the behavioral and lungs systems; whereas the minimum of shared chemicals was three (between the musculoskeletal system - vascular system, and the immunological system - nervous system associations; both having a low pS). Among the identified associations between systems, some are already known. For example, the high association between mutagenic and tumorigenic is coherent as it is widely accepted that tumorigenesis is a multistep process, depending on a sequential accumulation of mutations within tissue cells (Ashkenazi et al., 2008). Similarly, many chemicals that induce liver injury are also known to induce kidney injury, that is reflected by the liver-kidney association (Regner and Singbartl, 2016). Another strong association is observed between reproductive and specific developmental abnormal, which has already been demonstrated in various living organisms, including humans for some EDs (Krysiak-Baltyn et al., 2012)(Kinch et al., 2016)(Radke et al., 2018)(Ullah et al., 2019). In contrast, the association lungs-behavioral was more surprising, and no clear evidence was retrieved in the literature. That may be a

consequence of the broad definition of behavioral in RTECS, that consider for example anticonvulsant, sleep, euphoria, tremor, food intake, ataxia, analgesia, headache and alteration of classical conditioning.

System-system association network in the protein space (pS-SAN)

To explore potential mode of action of chemicals, and predict which systems may be affected by chemical exposure, a second model called pS-SAN (protein system-system associations network) was developed. This model is based on the known protein information related to the chemicals, and their associated systems. To create the model, each compound described in RTECS database was matched with the chemicals from STITCH database, that allowed extraction of protein information. From STITCH, a total of 7205 unique human proteins were compiled, for which 4735 unique compounds have been reported as being bioactive, for a total of 695,293 associations. However, some of these proteins might not be expressed in the system associated to the chemicals in RTECS. To limit the number of false positive protein-system associations, the protein expression in each tissue was collected from the HPA database, and only proteins showing moderate and high expression in a tissue correlated to one system described in RTECS were kept. As results, we obtained a total of 122,636 associations between 3162 proteins and 55 human tissues, that were mapped to the 24 systems from RTECS (Table 2). Therefore, the pS-SAN model was generated using 3162 human proteins bioactive to the 4735 selected chemicals and, known to be expressed (when information was available) in the 24 studies systems. The resulting pS-SAN model had 274 links between the 24 systems (Table S2). Based on the probability score ($pS > 0.07$), only 14 systems, among the 24, connected via 48 links had highly significant associations (Figure 3). The maximum of shared proteins was 2143 between the reproductive and

gastrointestinal systems, whereas the minimum of shared proteins is three as deciphered for associations between the musculoskeletal and nervous systems and, the cardiac and nutritional systems. Compare to the cS-SAN model, the interactions between the systems give different information. Strong associations are observed between different systems such as between endocrine-kidney, sense organs & special senses-blood, and kidney-gastrointestinal.

Deciphering novel links between endocrine disruptors and biological systems

The developed pS-SAN model was used to assess which system(s) can be affected by a chemical via protein complexes. As examples, we tested three EDs (AC, VA, and TCDD), to which protein bioactivities have been reported. For each ED, known chemical-protein links were extracted from the Comparative Toxicogenomics Database (CTD). Considering only information regarding curated data and human proteins, respectively 15,269, 7456 and 3985 proteins for VA, AC and TCDD were extracted. Each of the protein sets was independently screened against the pS-SAN model, to identify and, to prioritize the systems that may be significantly perturbed (Table 3).

Acetaminophen, the most commonly used analgesic drug in the world, was predicted to be highly connected to the liver, kidney, endocrine, lungs, and gastrointestinal systems. Several increasing concerns exist over long-term adverse effects from AC exposure. Recent studies have reported some evidence on renal, pulmonary, endocrine, neurobehavioral, skin reactions and cardiovascular toxicity (McCrae et al., 2018) (Kennon-McGill and McGill, 2018), such supporting the significant prediction from the pS-SAN model. Another study reported some evidence that mild analgesic can have endocrine disruptive properties able to alter

animal and human reproduction function (Kristensen et al., 2016). Regarding the medication valproic acid, significant associations with the brain, the endocrine, the kidney, the liver, the lungs and the peripheral nerve and sensation were retrieved. Looking at the literature, manifestations of VA toxicity such as cerebral edema, respiratory depression, metabolic and hematologic derangements and liver toxicity were reported, therefore supporting the pS-SAN predictions (Sztajnkrzyer, 2002). For the last screened chemical, the TCDD, associations to systems were less statistically significant i.e. kidney, specific developmental abnormalities, chronic data, biochemical and behavioral. Looking at the literature, in human, no clear evidence has been described so far. Nevertheless, some studies have shown some impact on the expression of some genes in kidney and peripheral blood abnormalities in mice (Fujisawa et al., 2018)(Wang et al., 2019). Looking at the compiled lists of proteins for the three EDs, a total of 17,160 unique proteins was retrieved, with disparate overlaps between the compounds (Figure 4). AC and VA had more common proteins than TCDD and VA/AC, that can be explained as these compounds belong to different chemical classes (drugs for VA and AC, pesticide for TCDD), and therefore were designed to have different modes of action.

Discussion

Network science has shown some benefit in human disease analysis, following the assumption that the functions of molecular components in a human cell are closely connected, and that a disease is rarely a consequence of a unique genetic variation, but a consequence of perturbations of complex intracellular and extracellular networks linking tissues and organ systems(Barabási et al., 2011). Such perturbations can be caused by diverse sources –not genetic- such as chemical exposure, physical stress, psychological

stress, etc. that are defined as the exposome, which corresponds to the totality of exposure over the lifetime. Concepts integrating both genetics and environmental factors will be beneficial for prediction of health and also therapies (Barouki et al., 2018). Toxicity can be presumed to be a linkage of perturbations at several layers of complexity. Integration of different types of large datasets into systems chemical toxicology, using network sciences, can help to explore and to assess chemical toxicities. Although some studies have demonstrated that the development of systems chemical toxicology models with the application of computational network biology may help in the understanding of chemical toxicity in human (Nie et al., 2015)(Hartung et al., 2017)(Taboureau and Audouze, 2016), the ability to assess chemical toxicity at an early stage remains a tremendous challenge for the regulatory agencies. New concepts and technologies are still needed to identify precise chemical health risk. The present study supports the concept of Integrated Approaches to Testing and Assessment (IATA) proposed by OECD [OECD (2017), *Guidance Document on the Reporting of Defined Approaches to be Used Within Integrated Approaches to Testing and Assessment*, OECD Series on Testing and Assessment, No. 255, OECD Publishing, Paris, (Sakuratani et al., 2018), that consists to integrate different types of data, and to perform hazard identification and safety assessment of chemicals(Casati, 2018).

The RTECS database allows to have access to a large set of chemicals with their assessment to a panel of toxicity endpoints. However, their mechanisms of action at the molecular level, and at the tissue level were not characterized. A chemical is more likely to cause toxicity in the organ/tissue where it is more likely to accumulate, and our study is a first step in a translational linkage among different data layers (proteins, tissues, systems). It allows to explore in a systematic way the effect of a chemical exposure in a human system. The proposed approach is highly depending on the aggregated data. Moreover, the curation of

these diverse sources of information is a key issue in the interpretation of the outcomes. We are aware that our toxicological system network has some limitations in term of chemical-protein bioactivities and protein-tissue annotations. To improve such predictive modeling, toxicokinetic information of a molecule, gene expression data of proteins and, pathways annotations could be other sources of knowledge to consider. Furthermore, one must take into consideration the so-called ‘Matthew effect’, that reflects the difference, in term of available documentation and scientific literature, between well investigated chemicals and less well studied environmental compounds(Grandjean et al., 2011). We can see such contrast with our case studies i.e. acetaminophen (largely studied, 28,340 publications in PubMed, May 14, 2020) and TCDD (less studied, 9939 publications in PubMed, May 14, 2020).


Conclusion

The ability to assess chemical toxicity at an early stage remains a tremendous challenge for the regulatory agencies. A better understanding of the effects for a molecule at a multiscale of the biological organization (cell, tissue, and organ) may lead to a better identification of chemical mode of action, allowing to offer a better knowledge for future drug development and for risk assessment of environmental pollutants. With the developed integrative toxicological systems network, possible linkage between chemicals and systems may be deciphered, allowing to suggest potential risk of toxicity. With the development of exposure science, the access of population-level observations and generation of IATA, new insights on human toxicity effects are expected to pave the way of chemical risk assessment.

Acknowledgements

The authors would like to acknowledge the OBERON project (<https://oberon-4eu.com/>), a project funded by the European Union's Horizon 2020 research and innovation program under grant agreement no. 825712. This work was also supported by the Université de Paris and the French National Institute of Health and Medical Research (Inserm).

Conflicts of Interest

The authors declare no conflict of interest. 

Author contributions

Conceived and designed the experiments: KA, OT. Performed the experiments: WE, KA, OT. Analyzed the data: KA, OT. Wrote the paper: KA, OT.

References

- Agier, L., Basagaña, X., Maitre, L., Granum, B., Bird, P.K., Casas, M., Ofstedal, B., Wright, J., Andrusaityte, S., de Castro, M., Cequier, E., Chatzi, L., Donaire-Gonzalez, D., Grazuleviciene, R., Haug, L.S., Sakhi, A.K., Leventakou, V., McEachan, R., Nieuwenhuijsen, M., Petraviciene, I., Robinson, O., Roumeliotaki, T., Sunyer, J., Tamayo-Uria, I., Thomsen, C., Urquiza, J., Valentin, A., Slama, R., Vrijheid, M., Siroux, V., 2019. Early-life exposome and lung function in children in Europe: an analysis of data from the longitudinal, population-based HELIX cohort. *Lancet Planet. Health* 3, e81–e92. [https://doi.org/10.1016/S2542-5196\(19\)30010-5](https://doi.org/10.1016/S2542-5196(19)30010-5)
- Aguayo-Orozco, A., Audouze, K., Siggaard, T., Barouki, R., Brunak, S., Taboureau, O., 2019. sAOP: Linking chemical stressors to Adverse Outcomes Pathway Networks. *Bioinforma. Oxf. Engl.* <https://doi.org/10.1093/bioinformatics/btz570>

365 Ankley, G.T., Bennett, R.S., Erickson, R.J., Hoff, D.J., Hornung, M.W., Johnson, R.D., Mount,
366 D.R., Nichols, J.W., Russom, C.L., Schmieder, P.K., Serrano, J.A., Tietge, J.E., Villeneuve, D.L.,
367 2010. Adverse outcome pathways: a conceptual framework to support ecotoxicology
368 research and risk assessment. *Environ. Toxicol. Chem.* 29, 730–741.
369 <https://doi.org/10.1002/etc.34>

370 Ashkenazi, R., Gentry, S.N., Jackson, T.L., 2008. Pathways to tumorigenesis--modeling
371 mutation acquisition in stem cells and their progeny. *Neoplasia N. Y. N* 10, 1170–1182.
372 <https://doi.org/10.1593/neo.08572>

373 Audouze, K., Juncker, A.S., Roque, F.J.S.S.A., Krysiak-Baltyn, K., Weinhold, N., Taboureau, O.,
374 Jensen, T.S., Brunak, S., 2010. Deciphering diseases and biological targets for environmental
375 chemicals using toxicogenomics networks. *PLoS Comput. Biol.* 6, e1000788.
376 <https://doi.org/10.1371/journal.pcbi.1000788>

377 Audouze, K., Sarigiannis, D., Alonso-Magdalena, P., Brochot, C., Casas, M., Vrijheid, M.,
378 Babin, P.J., Karakitsios, S., Coumoul, X., Barouki, R., 2020. Integrative Strategy of Testing
379 Systems for Identification of Endocrine Disruptors Inducing Metabolic Disorders-An
380 Introduction to the OBERON Project. *Int. J. Mol. Sci.* 21.
381 <https://doi.org/10.3390/ijms21082988>

382 Barabási, A.-L., Gulbahce, N., Loscalzo, J., 2011. Network medicine: a network-based
383 approach to human disease. *Nat. Rev. Genet.* 12, 56–68. <https://doi.org/10.1038/nrg2918>

384 Barouki, R., Audouze, K., Coumoul, X., Demenais, F., Gauguier, D., 2018. Integration of the
385 human exposome with the human genome to advance medicine. *Biochimie* 152, 155–158.
386 <https://doi.org/10.1016/j.biochi.2018.06.023>

387 Bopp, S.K., Kienzler, A., Richarz, A.-N., van der Linden, S.C., Paini, A., Parissis, N., Worth, A.P.,
388 2019. Regulatory assessment and risk management of chemical mixtures: challenges and

ways forward. *Crit. Rev. Toxicol.* 49, 174–189.

<https://doi.org/10.1080/10408444.2019.1579169>

Browne, P., Noyes, P.D., Casey, W.M., Dix, D.J., 2017. Application of Adverse Outcome Pathways to U.S. EPA’s Endocrine Disruptor Screening Program. *Environ. Health Perspect.* 125, 096001. <https://doi.org/10.1289/EHP1304>

Carvaillo, J.-C., Barouki, R., Coumoul, X., Audouze, K., 2019. Linking Bisphenol S to Adverse Outcome Pathways Using a Combined Text Mining and Systems Biology Approach. *Environ. Health Perspect.* 127, 47005. <https://doi.org/10.1289/EHP4200>

Casati, S., 2018. Integrated Approaches to Testing and Assessment. *Basic Clin. Pharmacol. Toxicol.* 123 Suppl 5, 51–55. <https://doi.org/10.1111/bcpt.13018>

Davis, A.P., Grondin, C.J., Johnson, R.J., Sciaky, D., King, B.L., McMorran, R., Wiegers, J., Wiegers, T.C., Mattingly, C.J., 2017. The Comparative Toxicogenomics Database: update 2017. *Nucleic Acids Res.* 45, D972–D978. <https://doi.org/10.1093/nar/gkw838>

Fujisawa, N., Yoshioka, W., Yanagisawa, H., Tohyama, C., 2018. Roles of cytosolic phospholipase A2 α in reproductive and systemic toxicities in 2,3,7,8-tetrachlorodibenzo-p-dioxin-exposed mice. *Arch. Toxicol.* 92, 789–801. <https://doi.org/10.1007/s00204-017-2081-z>

Grandjean, P., Eriksen, M.L., Ellegaard, O., Wallin, J.A., 2011. The Matthew effect in environmental science publication: a bibliometric analysis of chemical substances in journal articles. *Environ. Health Glob. Access Sci. Source* 10, 96. <https://doi.org/10.1186/1476-069X-10-96>

Hartung, T., FitzGerald, R.E., Jennings, P., Mirams, G.R., Peitsch, M.C., Rostami-Hodjegan, A., Shah, I., Wilks, M.F., Sturla, S.J., 2017. Systems Toxicology: Real World Applications and Opportunities. *Chem. Res. Toxicol.* 30, 870–882.

413 <https://doi.org/10.1021/acs.chemrestox.7b00003>

414 Jornod, F., Rugard, M., Tamisier, L., Coumoul, X., Andersen, H.R., Barouki, R., Audouze, K.,
415 2020. AOP4EUpest: Mapping of pesticides in Adverse Outcome Pathways using a text mining
416 tool. *Bioinformatics*, May 28:btaa545. <https://doi.org/10.1093/bioinformatics/btaa545>.

417 Judson, R.S., Martin, M.T., Egeghy, P., Gangwal, S., Reif, D.M., Kothiya, P., Wolf, M., Cathey,
418 T., Transue, T., Smith, D., Vail, J., Frame, A., Mosher, S., Cohen Hubal, E.A., Richard, A.M.,
419 2012. Aggregating data for computational toxicology applications: The U.S. Environmental
420 Protection Agency (EPA) Aggregated Computational Toxicology Resource (ACToR) System.
421 *Int. J. Mol. Sci.* 13, 1805–1831. <https://doi.org/10.3390/ijms13021805>

422 Kennon-McGill, S., McGill, M.R., 2018. Extrahepatic toxicity of acetaminophen: critical
423 evaluation of the evidence and proposed mechanisms. *J. Clin. Transl. Res.* 3, 297–310.

424 Kinch, C.D., Kurrasch, D.M., Habibi, H.R., 2016. Adverse morphological development in
425 embryonic zebrafish exposed to environmental concentrations of contaminants individually
426 and in mixture. *Aquat. Toxicol. Amst. Neth.* 175, 286–298.
427 <https://doi.org/10.1016/j.aquatox.2016.03.021>

428 Kitsak, M., Sharma, A., Menche, J., Guney, E., Ghiassian, S.D., Loscalzo, J., Barabási, A.-L.,
429 2016. Tissue Specificity of Human Disease Module. *Sci. Rep.* 6, 35241.
430 <https://doi.org/10.1038/srep35241>

431 Knapen, D., Vergauwen, L., Villeneuve, D.L., Ankley, G.T., 2015. The potential of AOP
432 networks for reproductive and developmental toxicity assay development. *Reprod. Toxicol.*
433 *Elmsford N* 56, 52–55. <https://doi.org/10.1016/j.reprotox.2015.04.003>

434 Knudsen, T., Martin, M., Chandler, K., Kleinstreuer, N., Judson, R., Sipes, N., 2013. Predictive
435 models and computational toxicology. *Methods Mol. Biol. Clifton NJ* 947, 343–374.
436 https://doi.org/10.1007/978-1-62703-131-8_26

437 Konkel, L., 2018. Reproductive Headache? Investigating Acetaminophen as a Potential
438 Endocrine Disruptor. *Environ. Health Perspect.* 126, 032001.
439 <https://doi.org/10.1289/EHP2478>

440 Krewski, D., Acosta, D., Andersen, M., Anderson, H., Bailer, J.C., Boekelheide, K., Brent, R.,
441 Charnley, G., Cheung, V.G., Green, S., Kelsey, K.T., Kerkvliet, N.I., Li, A.A., McCray, L., Meyer,
442 O., Patterson, R.D., Pennie, W., Scala, R.A., Solomon, G.M., Stephens, M., Yager, J., Zeise, L.,
443 2010. Toxicity testing in the 21st century: a vision and a strategy. *J. Toxicol. Environ. Health B*
444 *Crit. Rev.* 13, 51–138. <https://doi.org/10.1080/10937404.2010.483176>

445 Kristensen, D.M., Mazaud-Guittot, S., Gaudriault, P., Lesné, L., Serrano, T., Main, K.M., Jégou,
446 B., 2016. Analgesic use - prevalence, biomonitoring and endocrine and reproductive effects.
447 *Nat. Rev. Endocrinol.* 12, 381–393. <https://doi.org/10.1038/nrendo.2016.55>

448 Kristensen, D.M., Skalkam, M.L., Audouze, K., Lesné, L., Desdoits-Lethimonier, C.,
449 Frederiksen, H., Brunak, S., Skakkebaek, N.E., Jégou, B., Hansen, J.B., Junker, S., Leffers, H.,
450 2011. Many putative endocrine disruptors inhibit prostaglandin synthesis. *Environ. Health*
451 *Perspect.* 119, 534–541. <https://doi.org/10.1289/ehp.1002635>

452 Krysiak-Baltyn, K., Toppari, J., Skakkebaek, N.E., Jensen, T.S., Virtanen, H.E., Schramm, K.-W.,
453 Shen, H., Vartiainen, T., Kiviranta, H., Taboureau, O., Audouze, K., Brunak, S., Main, K.M.,
454 2012. Association between chemical pattern in breast milk and congenital cryptorchidism:
455 modelling of complex human exposures. *Int. J. Androl.* 35, 294–302.
456 <https://doi.org/10.1111/j.1365-2605.2012.01268.x>

457 Landrigan, P.J., Fuller, R., Acosta, N.J.R., Adeyi, O., Arnold, R., Basu, N.N., Baldé, A.B.,
458 Bertollini, R., Bose-O'Reilly, S., Boufford, J.I., Breyse, P.N., Chiles, T., Mahidol, C., Coll-Seck,
459 A.M., Cropper, M.L., Fobil, J., Fuster, V., Greenstone, M., Haines, A., Hanrahan, D., Hunter,
460 D., Khare, M., Krupnick, A., Lanphear, B., Lohani, B., Martin, K., Mathiasen, K.V., McTeer,

461 M.A., Murray, C.J.L., Ndahimananjara, J.D., Perera, F., Potočník, J., Preker, A.S., Ramesh, J.,
 462 Rockström, J., Salinas, C., Samson, L.D., Sandilya, K., Sly, P.D., Smith, K.R., Steiner, A.,
 463 Stewart, R.B., Suk, W.A., van Schayck, O.C.P., Yadama, G.N., Yumkella, K., Zhong, M., 2018.
 464 The Lancet Commission on pollution and health. *Lancet Lond. Engl.* 391, 462–512.
 465 [https://doi.org/10.1016/S0140-6736\(17\)32345-0](https://doi.org/10.1016/S0140-6736(17)32345-0)
 466 Lee, S., Lee, D.K., 2018. What is the proper way to apply the multiple comparison test?
 467 *Korean J. Anesthesiol.* 71, 353–360. <https://doi.org/10.4097/kja.d.18.00242>
 468 Leist, M., Ghallab, A., Graepel, R., Marchan, R., Hassan, R., Bennekou, S.H., Limonciel, A.,
 469 Vinken, M., Schildknecht, S., Waldmann, T., Danen, E., van Ravenzwaay, B., Kamp, H.,
 470 Gardner, I., Godoy, P., Bois, F.Y., Braeuning, A., Reif, R., Oesch, F., Drasdo, D., Höhme, S.,
 471 Schwarz, M., Hartung, T., Braunbeck, T., Beltman, J., Vrieling, H., Sanz, F., Forsby, A.,
 472 Gadaleta, D., Fisher, C., Kelm, J., Fluri, D., Ecker, G., Zdrazil, B., Terron, A., Jennings, P., van
 473 der Burg, B., Dooley, S., Meijer, A.H., Willighagen, E., Martens, M., Evelo, C., Mombelli, E.,
 474 Taboureau, O., Mantovani, A., Hardy, B., Koch, B., Escher, S., van Thriel, C., Cadenas, C.,
 475 Kroese, D., van de Water, B., Hengstler, J.G., 2017. Adverse outcome pathways:
 476 opportunities, limitations and open questions. *Arch. Toxicol.* 91, 3477–3505.
 477 <https://doi.org/10.1007/s00204-017-2045-3>
 478 McCrae, J.C., Morrison, E.E., MacIntyre, I.M., Dear, J.W., Webb, D.J., 2018. Long-term
 479 adverse effects of paracetamol - a review. *Br. J. Clin. Pharmacol.* 84, 2218–2230.
 480 <https://doi.org/10.1111/bcp.13656>
 481 Nie, W., Lv, Y., Yan, L., Chen, X., Lv, H., 2015. Prediction and Characterisation of the System
 482 Effects of Aristolochic Acid: A Novel Joint Network Analysis towards Therapeutic and
 483 Toxicological Mechanisms. *Sci. Rep.* 5, 17646. <https://doi.org/10.1038/srep17646>
 484 Pittman, M.E., Edwards, S.W., Ives, C., Mortensen, H.M., 2018. AOP-DB: A database resource

485 for the exploration of Adverse Outcome Pathways through integrated association networks.
486 Toxicol. Appl. Pharmacol. 343, 71–83. <https://doi.org/10.1016/j.taap.2018.02.006>
487 Radke, E.G., Braun, J.M., Meeker, J.D., Cooper, G.S., 2018. Phthalate exposure and male
488 reproductive outcomes: A systematic review of the human epidemiological evidence.
489 Environ. Int. 121, 764–793. <https://doi.org/10.1016/j.envint.2018.07.029>
490 Raghavan, R., Romano, M.E., Karagas, M.R., Penna, F.J., 2018. Pharmacologic and
491 Environmental Endocrine Disruptors in the Pathogenesis of Hypospadias: a Review. Curr.
492 Environ. Health Rep. 5, 499–511. <https://doi.org/10.1007/s40572-018-0214-z>
493 Regner, K.R., Singbartl, K., 2016. Kidney Injury in Liver Disease. Crit. Care Clin. 32, 343–355.
494 <https://doi.org/10.1016/j.ccc.2016.03.005>
495 Rugard, M., Coumoul, X., Carvaillo, J.-C., Barouki, R., Audouze, K., 2020. Deciphering Adverse
496 Outcome Pathway Network Linked to Bisphenol F Using Text Mining and Systems Toxicology
497 Approaches. Toxicol. Sci. Off. J. Soc. Toxicol. 173, 32–40.
498 <https://doi.org/10.1093/toxsci/kfz214>
499 Sakuratani, Y., Horie, M., Leinala, E., 2018. Integrated Approaches to Testing and
500 Assessment: OECD Activities on the Development and Use of Adverse Outcome Pathways
501 and Case Studies. Basic Clin. Pharmacol. Toxicol. 123 Suppl 5, 20–28.
502 <https://doi.org/10.1111/bcpt.12955>
503 Szklarczyk, D., Santos, A., von Mering, C., Jensen, L.J., Bork, P., Kuhn, M., 2016. STITCH 5:
504 augmenting protein-chemical interaction networks with tissue and affinity data. Nucleic
505 Acids Res. 44, D380–384. <https://doi.org/10.1093/nar/gkv1277>
506 Sztajnkrzyer, M.D., 2002. Valproic acid toxicity: overview and management. J. Toxicol. Clin.
507 Toxicol. 40, 789–801. <https://doi.org/10.1081/clt-120014645>
508 Taboureau, O., Audouze, K., 2016. Human Environmental Disease Network: A computational

509 model to assess toxicology of contaminants. ALTEX. <https://doi.org/10.14573/altex.1607201>

510 Thul, P.J., Åkesson, L., Wiking, M., Mahdessian, D., Geladaki, A., Ait Blal, H., Alm, T., Asplund,
511 A., Björk, L., Breckels, L.M., Bäckström, A., Danielsson, F., Fagerberg, L., Fall, J., Gatto, L.,
512 Gnann, C., Hober, S., Hjelmare, M., Johansson, F., Lee, S., Lindskog, C., Mulder, J., Mulvey,
513 C.M., Nilsson, P., Oksvold, P., Rockberg, J., Schutten, R., Schwenk, J.M., Sivertsson, Å.,
514 Sjöstedt, E., Skogs, M., Stadler, C., Sullivan, D.P., Tegel, H., Winsnes, C., Zhang, C., Zwahlen,
515 M., Mardinoglu, A., Pontén, F., von Feilitzen, K., Lilley, K.S., Uhlén, M., Lundberg, E., 2017. A
516 subcellular map of the human proteome. *Science* 356.
517 <https://doi.org/10.1126/science.aal3321>

518 Ullah, A., Pirzada, M., Jahan, S., Ullah, H., Razak, S., Rauf, N., Khan, M.J., Mahboob, S.Z.,
519 2019. Prenatal BPA and its analogs BPB, BPF, and BPS exposure and reproductive axis
520 function in the male offspring of Sprague Dawley rats. *Hum. Exp. Toxicol.* 38, 1344–1365.
521 <https://doi.org/10.1177/0960327119862335>

522 Vermeulen, R., Schymanski, E.L., Barabási, A.-L., Miller, G.W., 2020. The exposome and
523 health: Where chemistry meets biology. *Science* 367, 392–396.
524 <https://doi.org/10.1126/science.aay3164>

525 Verrotti, A., D'Egidio, C., Mohn, A., Coppola, G., Chiarelli, F., 2011. Weight gain following
526 treatment with valproic acid: pathogenetic mechanisms and clinical implications. *Obes. Rev.*
527 *Off. J. Int. Assoc. Study Obes.* 12, e32-43. <https://doi.org/10.1111/j.1467-789X.2010.00800.x>

528 Wang, L., Kumar, M., Deng, Q., Wang, X., Liu, M., Gong, Z., Zhang, S., Ma, X., Xu-Monette,
529 Z.Y., Xiao, M., Yi, Q., Young, K.H., Ramos, K.S., Li, Y., 2019. 2,3,7,8-Tetrachlorodibenzo-p-
530 dioxin (TCDD) induces peripheral blood abnormalities and plasma cell neoplasms resembling
531 multiple myeloma in mice. *Cancer Lett.* 440–441, 135–144.
532 <https://doi.org/10.1016/j.canlet.2018.10.009>

533 Wild, C.P., 2005. Complementing the genome with an “exposome”: the outstanding
534 challenge of environmental exposure measurement in molecular epidemiology. *Cancer*
535 *Epidemiol. Biomark. Prev. Publ. Am. Assoc. Cancer Res. Cosponsored Am. Soc. Prev. Oncol.*
536 *14*, 1847–1850. <https://doi.org/10.1158/1055-9965.EPI-05-0456>
537 Wu, Q., Achebouché, R., Audouze, K., 2020. Computational systems biology as an animal-
538 free approach to characterize toxicological effects of persistent organic pollutants. *ALTEX*.
539 <https://doi.org/10.14573/altex.1910161>
540 Yoshioka, W., Tohyama, C., 2019. Mechanisms of Developmental Toxicity of Dioxins and
541 Related Compounds. *Int. J. Mol. Sci.* *20*. <https://doi.org/10.3390/ijms20030617>
542
543

544 Tables

545 **Table 1. List of the human systems used to build the systems toxicological models.** In the

546 'comments' field are more detailed terms associated to 'system name'.

System name	Comments
Behavioral	Irritability, sleep, euphoria, hallucination...
Biochemical	Enzyme inhibition (esterase...), induction, change of level, effect on co-enzyme (vitamin B, folate...)
Blood	
Brain	Brain and coverings
Cardiac	
Chronic data	Changes in ovarian/prostate/testicular/uterine weight
Endocrine	Androgenic, estrogenic, diabetes mellitus...
Gastrointestinal	
Immunological incl. allergic	
Kidney	Kidney, ureter and bladder
Liver	
Lung	Lung, thorax and respiration
Musculoskeletal	
Mutagenic	
Negative mutagenic	
Nervous system	
Nutritional and gross metabolic	Dehydration, weight loss/gain
Peripheral nerve and sensation	
Reproductive	Effects on embryo or fetus, effects on fertility, effects on newborn, maternal effects, paternal effects
Sense organs and special senses	Nose, eye, ear and taste
Skin and appendages	Skin, hair, nails
Specific developmental abnormalities	
Tumorigenic	
Vascular	

547

548

549 **Table 2. Mapping of the systems to tissues.** Systems were extracted from the RTECS
550 database, and tissues from the HPA database. Correspondences were done manually, based
551 on both, databases and mining of the literature. ND (no data), meaning that no tissue
552 information was retrieved for the corresponding system.

Tissues	Systems	Tissues	Systems
adrenal gland	endocrine	placenta	reproductive
appendix	gastrointestinal	prostate	reproductive
bone marrow	immunological including allergic	rectum	gastrointestinal
breast	reproductive	retina	sense organs and special senses
bronchus	lungs	salivary gland	gastrointestinal
caudate	brain	seminal vesicle	reproductive
cerebellum	brain	skeletal muscle	musculoskeletal
cerebral cortex	brain	skin	skin and appendages
cervix, uterine	reproductive	skin 1	skin and appendages
colon	gastrointestinal	skin 2	skin and appendages
duodenum	gastrointestinal	small intestine	gastrointestinal
endometrium 1	reproductive	smooth muscle	peripheral nerve and sensation
endometrium 2	reproductive	soft tissue 1	peripheral nerve and sensation
epididymis	reproductive	soft tissue 2	peripheral nerve and sensation
esophagus	gastrointestinal	spleen	blood
eye	sense organs and special senses	stomach 1	gastrointestinal
fallopian tube	reproductive	stomach 2	gastrointestinal
gallbladder	kidney	testis	reproductive
hair	skin and appendages	thyroid gland	endocrine
heart muscle	cardiac	tonsil	sense organs and special senses
hippocampus	brain	urinary bladder	kidney
hypothalamus	brain	vagina	reproductive
kidney	kidney	n.d.	behavioral
lactating breast	reproductive	n.d.	biochemical

liver	liver	n.d.	chronic data
lung	lungs	n.d.	mutagenic
lymph node	blood	n.d.	negative mutagenic
nasopharynx	lungs	n.d.	nervous system
oral mucosa	gastrointestinal	n.d.	nutritional and gross metabolic
ovary	reproductive	n.d.	specific developmental abnormalities
pancreas	gastrointestinal	n.d.	tumorigenic
parathyroid gland	endocrine	n.d.	vascular
pituitary gland	endocrine		

553

554

555 **Table 3.** Chemical-system associations results based on pS-SAN model. In *italic*, systems not
556 related to tissues in HPA. NS (non-significant) after p-val correction (at 5%).

Systems name	AC	TCDD	VA
<i>Behavioral</i>	NS	0.025	NS
<i>Biochemical</i>	NS	0.0031	NS
Blood	2.29e-11	NS	9.81e-05
Brain	3.62e-09	NS	3.67e-20
Cardiac	1.01e-13	NS	1.85e-08
<i>Chronic Data</i>	NS	1.01e-06	NS
Endocrine	5.43e-22	NS	6.06e-16
Gastrointestinal	1.38e-20	NS	4.61e-07
Immunological Including Allergic	7.92e-10	NS	0.00036
Kidney	2.32e-33	0.00087	5.37e-18
Liver	6.73e-37	NS	3.69e-09
Lungs	2.09e-21	NS	5.17e-11
Musculoskeletal	1.54e-09	NS	1.18e-05
<i>Mutagenic</i>	NS	NS	NS

<i>Negative Mutagenic</i>	0.0022	NS	NS
<i>Nervous System</i>	NS	NS	NS
<i>Nutritional and Gross Metabolic</i>	NS	NS	NS
Peripheral Nerve and Sensation	9.50e-13	NS	9.81e-18
Reproductive	1.86e-13	NS	4.32e-09
Sense Organs and Special Senses	2.87e-11	NS	0.0027
Skin and Appendages	3.28e-14	NS	1.94e-07
<i>Specific Developmental Abnormalities</i>	NS	6.05e-06	NS
<i>Tumorigenic</i>	NS	NS	NS
<i>Vascular</i>	NS	NS	NS

557

558

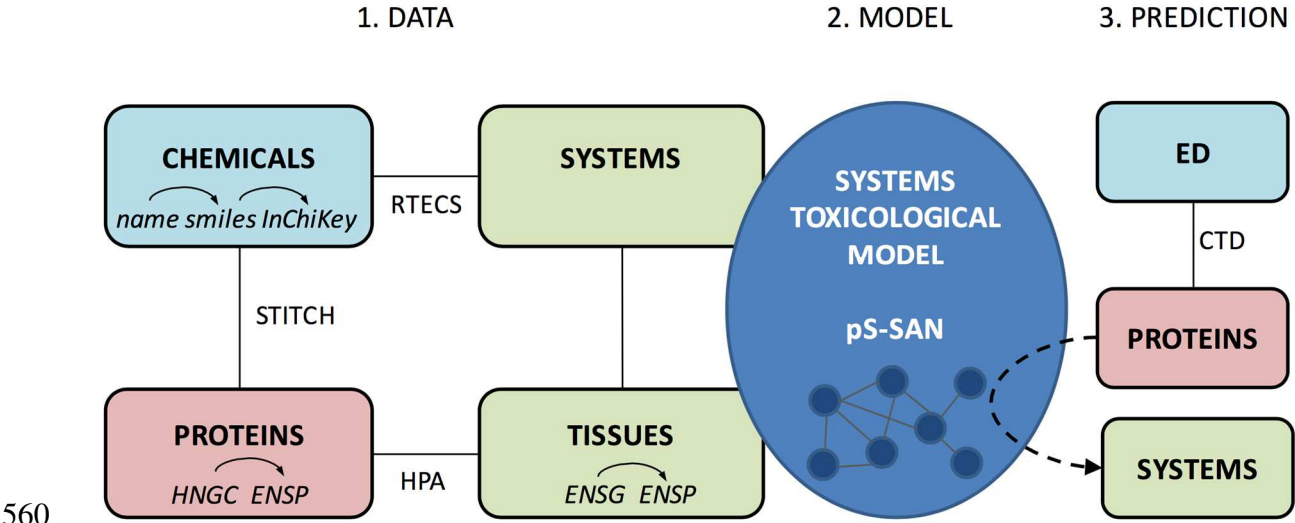


Figure 1: Workflow of the procedure implemented to develop the integrative toxicological network, that allow to predict biological systems affected by endocrine-disrupting chemicals. 1 Data: Systems that are known to be affected by environmental chemicals were extracted from the RTECS database, and proteins targeted by these chemicals were compiled from the STITCH database. Proteins were filtered using the Human Protein Atlas (HPA) database, to kept only the tissues and systems where proteins are known to be expressed. 2 Model: using the gathered chemical-protein-system information, an integrative systems toxicology model named pS-SAN (protein system-system association network) was performed. In this model, systems (blue circles) were associated by pairs based on their common proteins, that are affected by chemicals. 3. Prediction: the developed model was used to predict systems that can be affected by a compound, here an endocrine-disrupting chemical (ED), by statistical analysis on its known proteins from the Comparative Toxicogenomics Database (CTD).

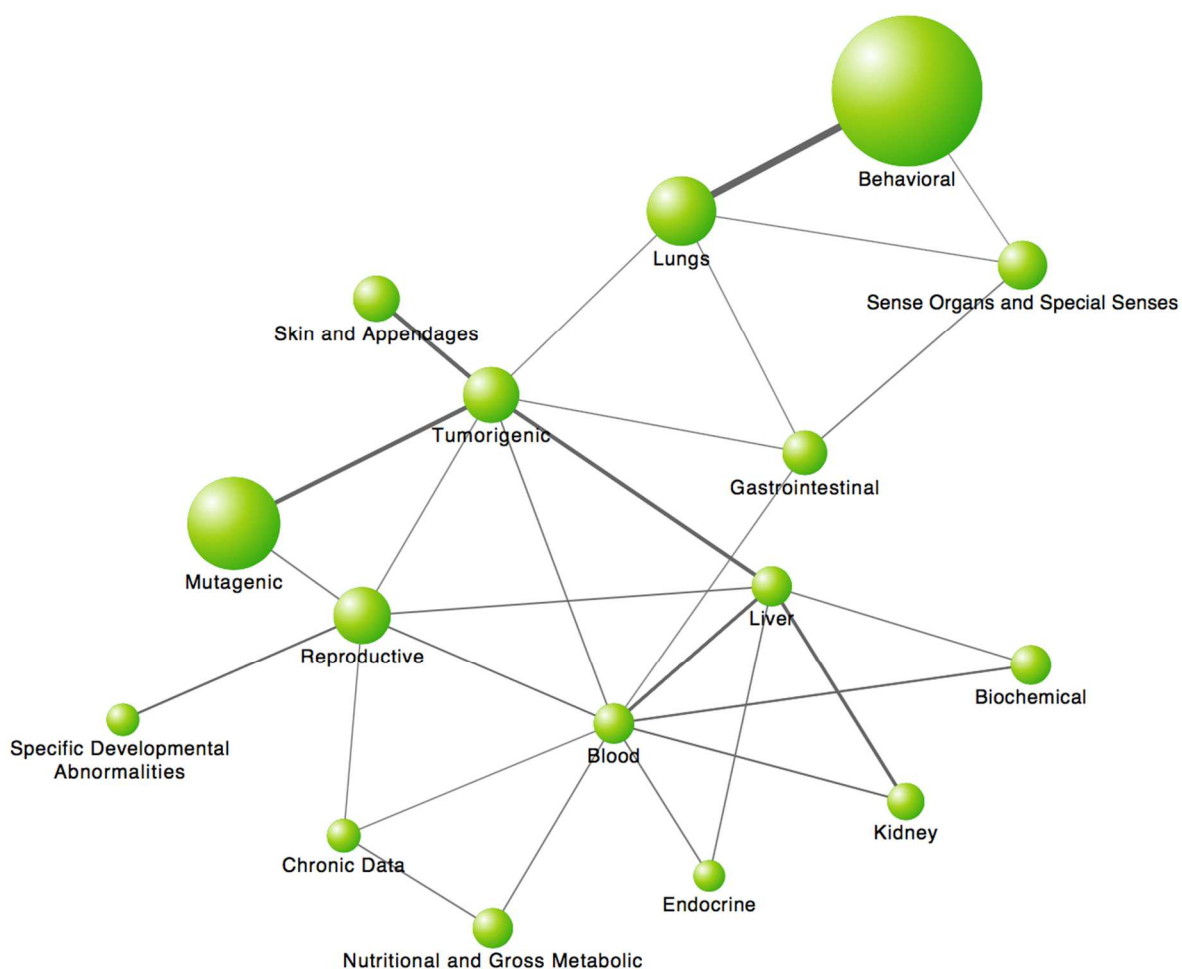


Figure 2: Representation of the top significant system-system associations network based on the chemical information (cS-SAN). Each node corresponds to a unique system. Systems are connected in the chemical space, meaning that two systems are associated if they shared at least one chemical. The edges represent the associations between the systems. The width of the edges is according to the calculated probabilistic scores (pS), which is proportional to the number of shared chemicals.

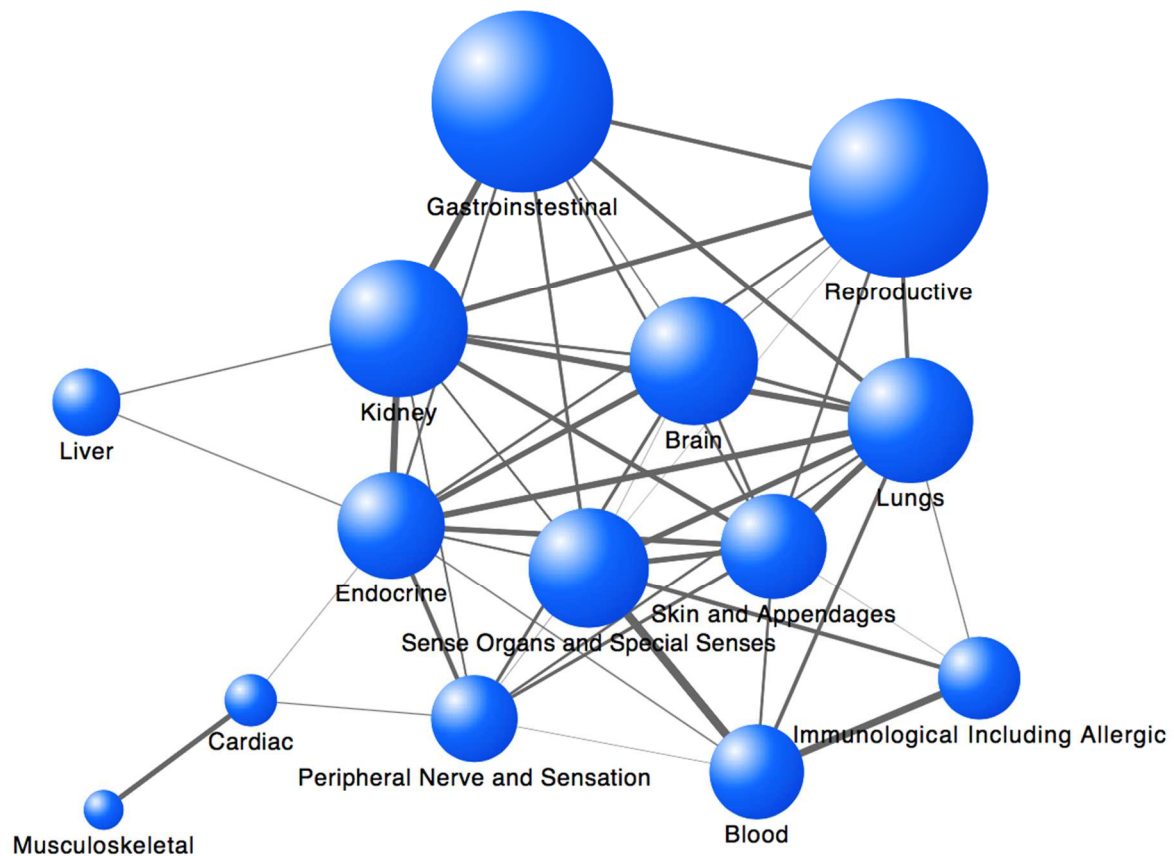
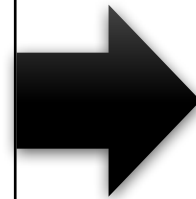
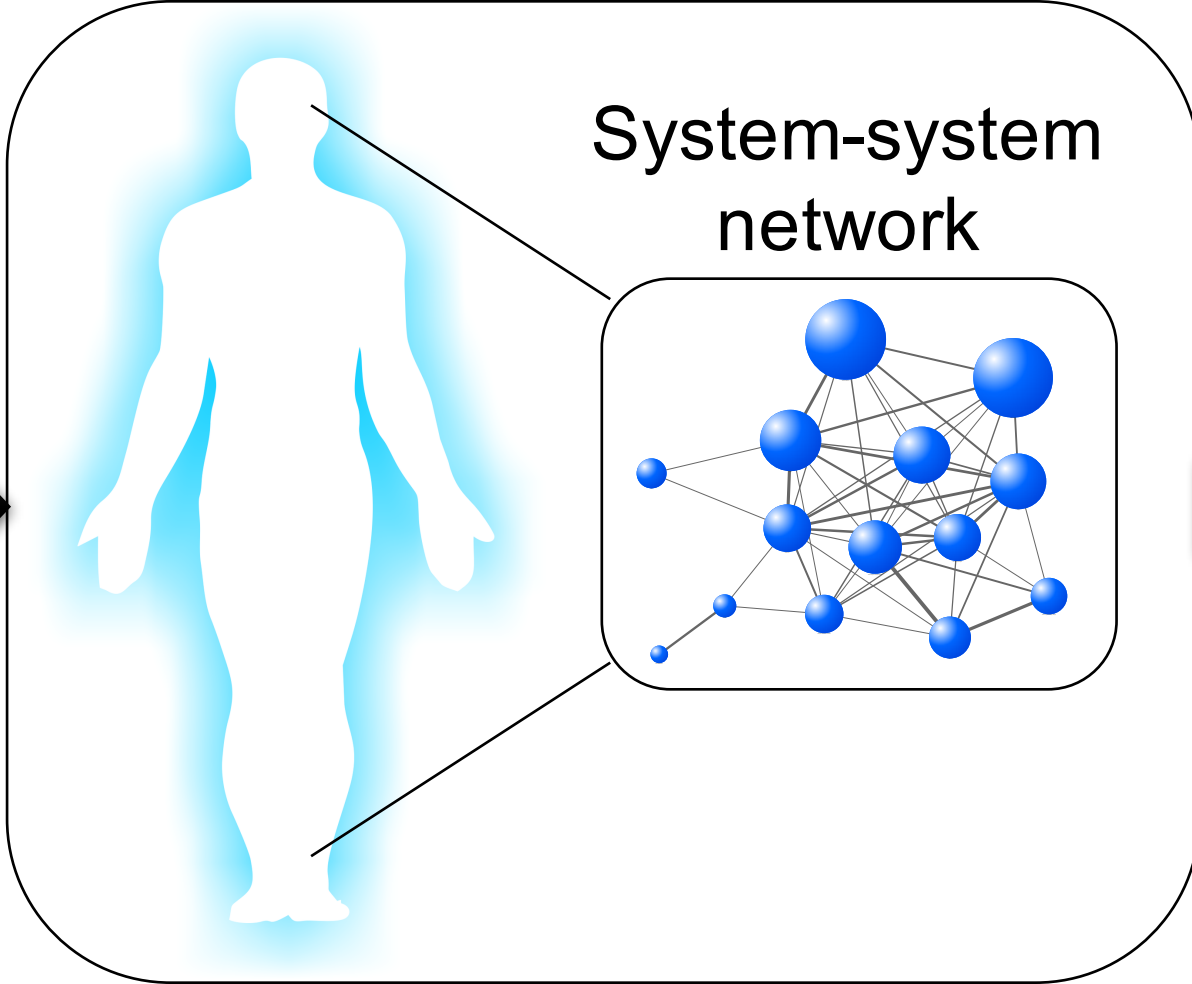
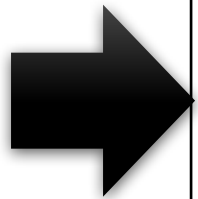


Figure 3: Representation of the top significant system-system associations network based on the protein information (pS-SAN). Each node corresponds to a unique system. Systems are connected in the protein space, meaning that two systems are associated if they shared at least one protein. The edges represent the associations between the systems. The width of the edges is according to the calculated probabilistic scores (pS), which is proportional to the number of common proteins.



Figure 4. Venn diagram showing the total number of proteins associated to the three endocrine-disrupting chemicals. The figure illustrates the number of proteins compiled from the CTD database, which is equivalent to the sum of values in each respective circle for each chemical. An overlapping shows the number of proteins commons between chemicals. The number at the center represents a total number of proteins targeted by all three compounds.

Environmental
chemical
exposure



Chemical
toxicity prediction